Elsevier Editorial System(tm) for LWT - Food Science and Technology
Manuscript Draft

Manuscript Number: LWT-D-13-00173

Title: Traditional Balsamic vinegar and Balsamic vinegar of Modena analyzed by Nuclear Magnetic Resonance spectroscopy coupled with multivariate data analysis

Article Type: Research Paper

Keywords: Balsamic Vinegar of Modena; Traditional Balsamic Vinegar of Modena; 1H NMR spectroscopy; PCA; GDA

Corresponding Author: Dr. Davide Bertelli,

Corresponding Author's Institution: University of Modena and Reggio Emilia

First Author: Davide Bertelli, dr.

Order of Authors: Davide Bertelli, dr.; Giulia Papotti; Riccardo Graziosi; Annalisa Maietti; Paola Tedeschi; Andrea Marchetti; Maria Plessi

Abstract: Balsamic vinegar of Modena (BVM) and traditional balsamic vinegar of Modena (TBVM) are highly appreciated typical Italian products. The quality control and authentication assurance of both these balsamic vinegars are very important topics. In the recent years, the interest to develop new and standardized analytical procedures, able to further enhance the quality and commercial value of these typical and unique products and to preserve them from possible sophistications and adulterations, is increased. In this work, 76 samples of both BVM and TBVM were analized by 1H NMR spectroscopy coupled with multivariate data analysis. The spectral data were analyzed by principal component analysis (PCA), general discriminant analysis (GDA) and classification tree analysis (CTA). The best and very promising model was obtained by a GDA which shows 98.6% of total variance explained by the first canonical function and a predictive capacity of 98.4% with a good separation between clusters. The signals of 5-HMF, □-glucopyranose, malic acid, succinic and acetic acids and a signal at 3.3 ppm referred to the glucose and fructose region were found to be the most statistically significant variables.

Suggested Reviewers: Augusta Caligiani
augusta.caligiani@unipr.it

Carmen Garcia-Parrilla
mcparrilla@us.es

Dirk W Lachenmeier
Lachenmeier@web.de

*Cover Letter

Università degli Studi di Modena e Reggio Emilia
DIPARTIMENTO DI SCIENZE DELLA VITA

Modena, 4 February , 2013

Dear **Katrin Hecht**
LWT- Food Science and Technology Editor

I enclose the manuscript "**Traditional Balsamic vinegar and Balsamic vinegar of Modena analyzed by Nuclear Magnetic Resonance spectroscopy coupled with multivariate data analysis**" and I would ask you to consider it with a view to publishing it in *LWT- Food Science and Technology.*

Corresponding author: Dr. Davide Bertelli, Dipartimento di Scienze della Vita, Università di Modena e Reggio Emilia, via Campi 183, 41125 Modena, Italy.
Tel.: +39 059 2055761; fax: +39 059 2055131.
E-mail address: davide.bertelli@unimore.it (D.Bertelli).

Best regards
Davide Bertelli

Recommended reviewers:

1. Prof.. Caligiani Augusta
   Dipartimento di Scienze degli Alimenti, Università di Parma, Parco Area delle Scienze, 95/A, 43124-Parma, Italy.
   E-mail address: augusta.caligiani@unipr.it.
2. Prof. M. Carmen Garcia-Parrilla
   Departamento Nutrición y Bromatología, Toxicología y Medicina Legal, Facultad de Farmacia, Universidad de Sevilla, 2, Sevilla 41012, Spain.
   e-mail: mcparrilla@us.es. Telephone 34954556760; fax 34954233765.
3. Dr. Dirk W. Lachenmeier

Chemisches und Veterinäruntersuchungsamt (CVUA) Karlsruhe, Weissenburger Strasse 3, D-76187 Karlsruhe, Germany.
E-mail: Lachenmeier@web.de Tel.: +49-721-926-5434. Fax: +49-721-926-5539.

***Highlights (for review)**

PGI BVM and PDO TBVM samples were analized by 1H NMR.

Spectral data were analyzed by PCA, GDA and CTA.

NMR and statistics seem to be a powerful tool in BVM and TBVM characterization.

A model able to classify the PGI and PDO samples was obtained

Traditional Balsamic vinegar and Balsamic vinegar of Modena analyzed by Nuclear Magnetic Resonance spectroscopy coupled with multivariate data analysis

(Giulia Papotti [a], Davide Bertelli [a*], Riccardo Graziosi [a], Annalisa Maietti [b], Paola Tedeschi [b], Andrea Marchetti [c], Maria Plessi[a])

[a] Dipartimento di Scienze della Vita, Università di Modena e Reggio Emilia, via Campi 183, 41125 Modena, Italy

[b] Dipartimento di Scienze Chimiche e Farmaceutiche, Università di Ferrara, via Fossato di Mortara 17/19, 44100 Ferrara, Italy

[c] Dipartimento di Scienze Chimiche e Geologiche, Università di Modena e Reggio Emilia, via Campi 183, 41125 Modena, Italy

∗Corresponding author: Dr. Davide Bertelli, Dipartimento di Scienze della Vita, Università di Modena e Reggio Emilia, via Campi 183, 41125 Modena, Italy.
Tel.: +39 059 2055761; fax: +39 059 2055131.
E-mail address: davide.bertelli@unimore.it (D.Bertelli).

Abstract

Balsamic vinegar of Modena (BVM) and traditional balsamic vinegar of Modena (TBVM) are highly appreciated typical Italian products. The quality control and authentication assurance of both these balsamic vinegars are very important topics. In the recent years, the interest to develop new and standardized analytical procedures, able to further enhance the quality and commercial value of these typical and unique products and to preserve them from possible sophistications and adulterations, is increased. In this work, 76 samples of both BVM and TBVM were analized by ^{1}H NMR spectroscopy coupled with multivariate data analysis. The spectral data were analyzed by principal component analysis (PCA), general discriminant analysis (GDA) and classification tree analysis (CTA). The best and very promising model was obtained by a GDA which shows 98.6% of total variance explained by the first canonical function and a predictive capacity of 98.4% with a good separation between clusters. The signals of 5-HMF, α-glucopyranose, malic acid, succinic and acetic acids and a signal at 3.3 ppm referred to the glucose and fructose region were found to be the most statistically significant variables.

KEYWORDS: Balsamic Vinegar of Modena , Traditional Balsamic Vinegar of Modena , ^{1}H NMR spectroscopy, PCA, GDA.

1. Introduction

The use of high resolution NMR in food authenticity is a subject of great interest in the scientific world, as long as, in the past few years, a widespread growth has been observed, in particular, in the usability of this technique as a fingerprint analysis tool coupled with multivariate data analysis (Bertelli, Papotti, Bortolotti, Marcazzan & Plessi, 2012; Bertelli et al., 2010; Mannina et al., 2012; Lopez-Rituerto et al., 2012; Koda, Furihata, Wei, Miyakawa & Tanokura, 2012; Krishnan, Kruger & Ratcliffe, 2005). The application of NMR in food characterization and control is nowadays very powerfull and outstanding due to several good reasons. This technique is well suited to analyze samples without any manipulation, so that, in most cases, extraction and purification are unnecessary, and it is not a destructive analysis. Owing to these advantages, this technique has attended a great development in food science, mainly concerned with the qualitative interpretation of the NMR spectra. This is particularly true for liquid food, such as wine (Papotti et al., 2012), fruit juice (Clausen, Pedersen, Bertram & Kidmose, 2011), beverages (Lachenmeier et al., 2005; Maes, Monakhova, Kuballa, Reusch & Lachenmeier, 2012) and vinegar (Dell'Oro, Ciambotti & Tsolakis, 2012; Van-Diep et al., 2011; Thomas & Jamin, 2009; Boffo, Tavares, Ferreira & Ferreira, 2009; Caligiani, Acquotti, Palla & Bocchi, 2007; Consonni & Gatti, 2004), which may be directly analyzed, after the addition of a deuterated solvent and an internal standard. Other advantages of NMR are the relatively easy and rapid acquisition of data (few minutes are required to acquire a simple ^{1}H NMR spectrum), the remarkable selectivity and identification of unknown compounds at a molecular level with high reproducibility and repeatability, the ability to furnish structural and quantitative information on a wide range of chemical species in a single NMR experiment, as long as, it is considered as a fingerprint method for authentication analysis.

BVM and TBVM are considered typical and well known Italian products, highly appreciated all over the world. Although BVM and TBVM have some organoleptic characteristics in common, they are different products. TBVM is a Protected Designation of Origin (PDO) product (Reg. CE n. 813/2000 April 17, 2000), owing to its typical production procedure and the well-defined area of origin. Nowadays, local vinegar houses, often founded by small family-run business, produce the TVBM, according to the ancient methods of production, whose origins are to be found in the Modenese traditions. Long fermentation and aging procedures, which require expertise and caution in respect of the maturation state of the product, contribute to develop the unique and unmistakable characteristics that we recognize today in this very valuable product. BVM has recently obtained the registration with Protected Geographical Indication (PGI) status, granted by the European Union (Reg. CE n. 583/2009 July 3th, 2009). They are both obtained from the alcoholic and acetic fermentation of cooked and concentrated grape musts, and this is the main characteristic that distinguishes balsamic vinegars from other vinegars, which are generally produced from alcoholic solution. TBVM and BVM mainly differ in the aging process and the production procedures. TBVM is aged in characteristic wooden barrels and may be found on the market in two different products according to the aging process: old (>12 and <25 years) and extra old (>25 years). It is allowed the inoculation of colonies of acetic bacteria, while the use of any extra additive is forbidden. BVM is a cheaper product, with maturation in wooden barrels from two months up to 3 years, and it is allowed to add vinegar obtained by wine acetification (10% v/v minimum) and caramel (2% v/v maximum) for color correction (D.M., December 3th, 1965). From the regulatory point of view, the quality control and authentication assurance of both these balsamic vinegars was tested by means of sensorial analysis and by very simple chemical–physical property determinations, like total acidity, density and dry residual. Here arises the interest to develop new and standardized analytical procedures, able to further enhance the

quality and commercial value of these typical and unique products and to preserve them from possible sophistications and adulterations. These new approaches, coupled with chemometric approaches, may provide helpful classification models for authentication and commercial quality characterization.

Several analytical studies focused their attention on the balsamic vinegars characterization and, some of them aim to investigate the modifications occurring during aging and the age determination (Cocchi, Lambertini, Mancini, Marchetti & Ulrici, 2002; Plessi, Bertelli & Miglietta, 2006; Plessi, Monzani & Coppini, 1989; Del Signore, Stancher & Calabrese, 2000; Theobald, Muller & Anklam, 1998; Antonelli, Chinnici & Masino, 2004; Chiavaro, Caligiani & Palla, 1998; Cocchi et al., 2006; Gullo, C.; Caggia, De Vero & Giudici, 2006). However, they are often time consuming approaches, not compatible with routine analyses. Among many advantages that the ^{1}H NMR spectroscopy offers, it may simultaneously determine the different metabolites of vinegar in few minutes, as required for food authenticity and quality control. To our knowledge, only few studies have been carried out on Traditional Balsamic Vinegar of Modena (TBVM) and Balsamic Vinegar of Modena (BVM), regarding their quality evaluation and valorization, using NMR (Cirlini, Caligiani & Palla, 2009; Consonni et al., 2008; Consonni, Cagliani & Rinaldini, 2008; Consonni & Cagliani, 2007). Here a characterization of both BVM and TBVM using high-resolution ^{1}H NMR spectroscopy coupled with multivariate statistical data analysis is presented. Besides, we aim to build a discriminant model able to characterize TBVM according to the aging process, which is actually the most required information for quality assessment, and, nowadays, no objective analytical techniques have been officially defined. The particular climatic characteristics, the soil and the grape varieties typically grown in Modena strongly contribute to make the TBVM a unique and unmistakable product. These characteristics and the particular production procedures used by the local producers, which carry on the Modenese traditions through the

art of cooking musts and drawing and topping up procedures among the wooden casks, make it is difficult to obtain statistical models which are representative of the intrinsic variability and peculiarities of TBVM. In this work, we applied supervised pattern recognition procedures never used before, and a very significant number of samples is available, that is non-obvoius, also considering the high price of the samples.

2. Materials and methods

2.1. Materials and sample preparation

A total of 76 samples of both TBVM and BVM have been analyzed. Among them, 23 were extra old (>25 years of aging) TBVM, 17 were old (>12 and <25 years of aging) TBVM and 36 were BVM of unknown aging (**Table 1**). All the TBVM and several BVM were provided by local vinegar houses, while the other BVM were purchased on the market. All TBVMs and BVMs are labelled as PDO and PGI products respectively. Samples were prepared into the Wilmad NMR tube (5 mm, Ultra-Imperial grade, 7 in. L, 526-PP, Sigma-Aldrich, Milan, Italy) by dissolving 0.1 g exactly weighed of vinegar into 500 μL of dimethyl sulphoxide-d_6 (DMSO-d_6) (Sigma-Aldrich, Milan, Italy), and 20 μL of tetramethylsilane (TMS) was added as reference compound. Standard compounds for metabolite assignments were from Sigma-Aldrich (Milan, Italy).

2.2. Physical and chemical determinations

Undiluted samples were used for °Brix measures, which were carried out with refractometer. Total acidity was determined by the titration Official Method (Resolution OIV-OENO 52-2000). *R ratio*, which is the rate between °Brix and Total Acidity % (p/v), and indicates the balance among sweet and sour tones, correlating in this way the density with the acidity (Gullo & De Vero, 2004; Satrioni, 2010), was also calculated for TBVM samples. This

parameter has always been used as a tool that, if properly interpreted, allowed to conduct correctly vinegar houses and to identify possible outliers, before performing NMR analysis.

2.3. NMR spectroscopy

To characterize samples ^{1}H NMR, ^{13}C NMR, bidimensional ^{1}H-^{13}C heteronuclear multiple-bond correlation (HMBC) and ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) spectra were acquired with a Bruker FT-NMR Avance 400 spectrometer (Bruker Biospin GmbH Rheinstetten, Karlsruhe, Germany) operating at 400.13 MHz for ^{1}H. All of the experiments were performed at 300 K and nonspinning. ^{1}H NMR data were acquired using the Bruker spin−echo sequence "cpmg1d" (Carr−Purcell−Meiboom−Gill, Bruker Library) without water presaturation. Acquisition parameters were as follows: time domain (number of data points), 32K; dummy scans, 4; acquisition time, 3.4210 s; delay time, 3.0 s; number of scans, 64; spectral width, 4789.27 Hz; fidres, 0.1461. Total acquisition time was 7 min and 46 s. The assignments have been carried out on the basis of the ^{13}C NMR, HMBC and HSQC analyses. The acquisition parameters of the ^{13}C NMR experiments were as follows: number of scans, 8K; dummy scans, 4; time domain (number 142 of data points), 32K; spectral width, 22075.055 Hz; acquisition time, 0.7422 s; delay time, 1.5 s; fidres, 0.6737 Hz. Total acquisition time was 5 h, 14 min, and 59 s. The acquisition parameters of the HMBC experiments were as follows: number of scans, 32; dummy scans, 16; time domain, 3K in the acquisition or direct HMBC dimension F2 (^{1}H) and 100 in indirect HMBC dimension F1 (^{13}C); spectral width, 5592.841 Hz in F2 (^{1}H) and 20124.465 Hz in F1 (^{13}C); digital resolution, 1.8206 Hz in F2 (^{1}H) and 201.245 Hz in F1 (^{13}C); acquisition time, 0.2747 s; delay time, 0.5 s; HMBC delay time, 62.5 ms. Total acquisition time was 82 min and 11 s. The acquisition parameters of the HSQC experiments were as follows: number of scans, 4; dummy scans, 12; time domain, 1K in the acquisition or direct HSQC dimension F2 (^{1}H) and 256 in indirect HSQC dimension F1 (^{13}C); spectral width, 5995.204 in F2 (^{1}H) and 19118.721

F1 (^{13}C); digital resolution, 5.855 Hz in F2 (^{1}H) and 74.682 Hz in F1 (^{13}C); acquisition time, 0.0854 s; delay time, 1.5 s. Total acquisition time was 27 min and 50 s. The chemical shifts were reported as δ_H (ppm) relative to TMS.

2.4. Spectral calculation

The classification models were obtained using the ^{1}H NMR spectra as intensity. The application of the ^{1}H NMR technique to vinegar samples generates very complicated spectra that need to be previously processed and subsequently analyzed by chemometric methods. Each spectrum generated a 16K data points corresponding to time domain, that is the number of points acquired and digitalized by the instrument along the spectral width, and then converted into a frequency domain spectrum by Fourier transform; these files were collected in a data set consisting of 16K variables and 76 samples. All ^{1}H NMR spectra were phased and calibrated using the TMS signal by the XWinNMR software package (Bruker Biospin GmbH Rheinstetten). To reduce the inhomogeneous proton NMR chemical shift, all spectra were aligned using the toolbox Icoshift 1.0 for MATLAB (Mathworks Inc., Natick, MA, USA) (Savorani, Tomasi & Engelsen, 2009). Finally, the spectra were baseline corrected by PLS_Toolbox version 5.2.2 for use with MATLAB (eigenvector Research Inc., Wenatchee, WA, USA). All the spectral regions devoid of signals and the residual solvent (DMSO-d_6) signal (region from 2.45 to 2.55 ppm) were not considered. The resulting data set refers to the complete spectral region (12362 variables). Two other data sets have been prepared, the first one referred to the low-frequency spectral region between 0.65 and 2.70 ppm (2586 variables), which principally contains the signals of acidic and aliphatic compounds, and the second one, which contains the signals of the midfrequency region, between 2.70 and 5.50 ppm (3499 variables).

2.5. Statistical analysis

Before the spectral analyses, all data were normalized, mean-centered, and scaled by the pareto-scaling method (Winning et al., 2009). To achieve a reliable classification, unsupervised and supervised pattern recognition procedures were applied to the data sets. Principal component analysis (PCA) was performed to verify the intrinsic variation in the data sets. Factor analysis (FA) (Burt, 1950) and general discriminant analysis (GDA) (McLachlan, 1992) were used to classify the vinegars according to their NMR fingerprint. To perform GDA, a reduction in variables with respect to complete data sets was necessary. For the complete spectral region data set the number of variables was reduced considering only the signals which presented a factorial weight during FA > |0.8|, and the resulting data set with 1453 variables and 76 samples was obtained. For the other two data sets, a less severe variables reduction was applied, by simply reducing in spectral resolution. The number of variables in fact was halved, compared to the original data sets, thus ensuring that each peak mantains its shape. After the construction of the models, to evaluate the classification performance, the leave-one out method was used as a validation procedure (Henrion & Henrion, 1994). The most significant signals resulted from GDA were integrated, using the software Amix 3.7.10 (Bruker Biospin GMBH, Rheinstetten, Germany), and used for classification trees analysis (CTA). The aim was to build a discriminant model able to fit adequately the aging process and the type of sample, in order to find a separation between the extra old TBVM and the old TBVM and identify the signals that allow such separation.

CTA is used to predict membership of cases or objects in the classes of a categorical dependent variable from their measurements on one or more predictor variables, and, in the past years, it has been successfully used in different areas of healthcare (Harper & Shahani, 2002; Ridley et al., 1998; Harper & Winslett, 2004), and food science (Bertelli, Plessi, Sabatini, Lolli & Grillenzoni, 2007; Cirlini, Caligiani, Palla & Palla, 2010). Three different building tree methods were applied to datasets: (a) Discriminantbased Univariate Splits

(DUS); (b) Discriminant-based Linear Combination Splits (DLCS); (c) Classification & Regression Tree-style Exhaustive Search for Univariate splits (C&RT). For all of these three building methods, the FACT-style direct stopping was used as stopping rule (Loh & Vanichestakul, 1988). To estimate the prediction capacity of the models, a one-third cross-validation method was applied.

All calculations were performed using the PLS_Toolbox version 5.2.2 for Matlab, Statistica 6.1 (StatSoft® Italia, Vigonza, Italy) and SPSS 13.0 (SPSS Inc., Chicago, IL, USA).

3. Results and discussion

One of the main differences that BVM and TBVM show at macroscopic level is the sugar concentration, which directly influences the density (°Brix). As well-known (Masino, Chinnici, Franchini, Ulrici & Antonelli, 2005), this parameter shows an increase with aging, and this is confirmed also by the density values measured in our samples, reported in **Table 1**. Another important parameter, which correlates density with total acidity, is *R ratio* (Gullo et al., 2004; Satrioni, 2010). For traditional balsamic vinegars produced according to set rules, the optimal value of *R ratio* is between 7–10 (Gullo et al., 2004). The 76 samples show *R* values close to this range, between 8.75 and 12.91 (**Table 1**). Although our results are slightly higher, all the TBVM samples were considered valid and used in the following analyses, since similar *R* values were obtained also by other authors (Consonni, Cagliani, Rinaldini & Incerti, 2008).

A ^{1}H NMR spectrum recorded at 400 MHz of a TBVM sample, with expansions of aliphatic/alcoholic, and sugar regions, is shown in **Figure 1**.

The sequence allows to suppress broad signals, in our case the water signal, which may be selectively removed, and to enhance narrow resonances. Unfortunately also the large signals of the hydroxyl groups are suppressed; anyway the spectra are very rich in information. This

sequence thus provides a sensitive means of investigating the composition of the balsamic vinegar samples object of this study. The metabolites were assigned on the basis of additional 2D NMR experiments of samples, by recording NMR spectra of pure compounds, and confirmed by comparing our results with literature data (Cirlini et al., 2009; Consonni, Cagliani, Rinaldini & Incerti, 2008; Consonni et al., 2008; Consonni et al., 2007). The principal metabolites assignments are summarized in **Table 2**. The simple comparison of the calibrated and normalized spectra (**Figure 2**) confirmed that BVM , TBVM old and TBVM extra old show different spectral characteristics, certainly due to the different production techniques, therefore to the consequent different composition. In the case of TBVMs, also the aging process may probably influence the spectral aspect of these samples. The main differences in the spectral intensities distribution are related to the signals of acetic acid, ethanol and acetoin. In TBVM, these compounds seem to be less concentrated in the majority of more aged samples as reported also by other authors. Besides, 5-HMF, fructose and glucose, 6-acetyl glucose and 2,3-butanediol were subjected to larger increase. These compounds have been described in the literature as the most significant ones for monitoring the aging process in TBVs (Caligiani et al., 2007; Masino et al., 2005; Theobald et al., 1998; Cirlini et al., 2009; Consonni et al., 2008). When a comparison between balsamic and traditional balsamic vinegars is performed, it is necessary to specify that decreases and increases in metabolite concentrations are not closely related to the aging process, therefore, it is more cautious, and more correct, to affirm that these differences are probably more related to the intrinsic diversity of products.

The PCA was performed on the ^{1}H NMR complete spectral region to check possible sample grouping (**Figure 3**). This model resulted in 8 PCs explaining 96.91% of the total variance, and it was able to discriminate only the BVMs from TBVMs, without any differentiation between old and extra old TBVMs. This result demonstrated the considerable complexity of

the system; therefore, in order to give a clearer interpretation of PCA model, and to verify whether the application of a supervised multivariate statistical analysis was able to classify the samples, GDA was applied. The analysis was performed on complete and also on expanded spectral region data sets, since a less severe variables reduction was necessary in the aliphatic/alcoholic and sugar regions. The best results were obtained using the forward stepwise procedure. The GDA model obtained from the complete spectra was able to group the samples in three evident clusters, corresponding to the type of sample (**Figure 4**). The first two canonical discriminant function (DF) explains 98.6% of the total variance, and the results of the leave-one out cross-validation show a predictive capacity of 98.4%. The first two DFs are particularly correlated with the signals of 5-HMF (C1H, C3H, C4H), α-glucopyranose (αC1H), malic acid (C3H), a signal at 3.3 ppm referred to the glucose and fructose region, succinic and acetic acids. These results were confirmed by the models obtained by analyzing the expanded spectral regions, and were both able to group the vinegars in three evident clusters (score plots not shown). The first DFs of aliphatic/alcoholic and sugar regions data sets explain 98.3% and 89.7% of the total variance respectively, and the results of the leave-one out cross-validation show a predictive capacity of 90.5% and 96.8% respectively. The first two DFs of each model are particularly correlated with the signals of succinic acid, malic acid (C2H), 6-acetyl glucose and the unknown signal at 1.74 ppm for the aliphatic/alcoholic region data set, α-glucopyranose (αC1H), 5-HMF ($C6H_2$), tartaric acid, β-glucopyranose (βC1H), β-fructopyranose (C6H), 2 signals at 3.3 ppm and 3.03 ppm referred to the glucose and fructose region for the sugar region data set. The signals, corresponding to the most significant compounds in GDA, were integrated and used to build a new CTA discriminant model, in order to evaluate their effects on the balsamic vinegars discrimination. The obtained results are summarized in **Table 3**. In general, when the results of CTA were judged, the number of terminal nodes must be considered and a tree with a small number of terminal

nodes must be preferred if the same capacity of classification is reached by different approaches. Considering the classification results and the number of terminal nodes, C&RT seems to be a good compromise between classification capacity and the complexity of the tree. In **Figure 5** the best tree obtained by C&RT method is reported. As evident, the model is able to discriminate between the extra old TBVM and the old TBVM and to identify the signals that allow such separation. The most interesting finding resulted from CTA is that in this model the discrimination between old TBVM and extra old TBVM is essentially due to the signals of acetic acid, 5-HMF (C4H), malic acid (C3H), β-glucopyranose (βC1H), corresponding to those compounds already identify as significant ones also in the literature (Caligiani et al., 2007; Consonni et al., 2004; Theobald et al., 1998; Consonni et al., 2008), and to the unknown compound at 1.74 ppm.

4. Conclusion

The NMR spectroscopy, coupled with multivariate analysis, has demonstrated to be a powerful tool in BVM and TBVM characterization and quality control. Statistical analysis showed that the signals of 5-HMF, α and β-glucopyranose, malic, succinic, tartaric and acetic acids, 6-acetyl glucose, a signal at 3.3 ppm referred to the glucose and fructose region were the most statistically significant variables. All these compounds are described in the literature as relevant ones for discriminating the balsamic vinegars and for monitoring the aging process. Theobald et al. (1998) established that the highest concentration of 5-HMF was found in TBV samples. The changes in sugars and in acetic acid contents occuring during the aging process of TBVs are a well known finding (Caligiani et al., 2007; Consonni et al., 2008). In 2004, as a part of NMR studies on Italian balsamic and traditional balsamic vinegars, Consonni et al. (2004) confirmed that, among several compounds, the highest content of malic acid was found for the older BV and TBV, while much lower values were found for the younger BVs.

Cocchi et al. (2002) proved that succinic acid increases in the young vinegars and decreases in the old ones. The hypothesis of formation of glucose and fructose acetates during maturation and aging of balsamic vinegar was verified by Cirlini et al. in 2009.

All these findings provide interesting additional knowledge on TBVs, usable to reinforce and safeguard the wealth of Modenese traditions, represented by these unique products.

Acknowledgements

The authors express their gratitude towards the "Assessore all'Agricoltura e alla Qualità del Territorio rurale" of the province of Modena for the financial efforts provided during our research. We also express our thanks to the stuff of C.I.G.S. (Centro Interdipartimentale Grandi Strumenti, Modena, Italy) for assistance during the experimental work and the Fondazione Cassa di Risparmio di Modena, for financial support given towards the purchase of the Bruker Avance 400 spectrometer.

References

Antonelli, A., Chinnici, F., & Masino, F. (2004). Heat-induced chemical modification of grape must as related to its concentration during the production of traditional balsamic vinegar: a preliminary approach. *Food Chemistry*, *88*, 63–68.

Bertelli, D., Lolli, M., Papotti, G., Bortolotti, L., Serra, G., & Plessi, M. (2010). Detection of honey adulteration by sugar syrups using one-dimensional and two-dimensional high-resolution nuclear magnetic resonance. *Journal of Agricultural and Food Chemistry*, *58*, 8495−8501.

Bertelli, D., Papotti, G., Bortolotti, L., Marcazzan, G. L., & Plessi, M. (2012). ^{1}H-NMR simultaneous identification of health-relevant compounds in propolis extracts. *Phytochemical Analysis*, *23*, 260−266.

Bertelli, D., Plessi, M., Sabatini, A.G., Lolli, M., & Grillenzoni, F. (2007). Classification of Italian honeys by mid-infrared diffuse reflectance spectroscopy (DRIFTS). *Food Chemistry, 101,* 1565-1570.

Boffo, E.F., Tavares, L.A., Ferreira, M.M.C., & Ferreira, A.G. (2009). Classification of Brazilian vinegars according to their ^{1}H NMR spectra by pattern recognition analysis. *LWT--Food Science and Technology, 42,* 1455-1460.

Burt, C. (1950). The factorial analysis of qualitative data. *British Journal of Psychology, 3,* 166–185.

Caligiani, A., Acquotti, D., Palla, G., & Bocchi, V. (2007). Identification and quantification of the main organic components of vinegars by high resolution ^{1}H NMR spectroscopy. *Analytica Chimica Acta, 585,* 110–119.

Chiavaro, E., Caligiani, A., & Palla, G. (1998). Chiral indicators of ageing in balsamic vinegars of Modena. *Italian Journal of Food Science, 4,* 329–337.

Cirlini, M., Caligiani, A., & Palla, G. (2009).Formation of glucose and fructose acetates during maturation and ageing of balsamic vinegars. *Food Chemistry, 112,* 51–56.

Cirlini, M., Caligiani, A., Palla, L., & Palla, G. (2010). HS-SPME/GC-MS and chemometrics for the classification of Balsamic Vinegars of Modena of different maturation and aging. *Food Chemistry, 124,* 1678-1683.

Clausen, M. R., Pedersen, B. H., Bertram, H. C., & Kidmose, U. (2011). Quality of sour cherry juice of different clones and cultivars (*Prunus cerasus L.*) determined by a combined sensory and NMR spectroscopic approach source. *Journal of Agricultural and Food Chemistry, 59,* 12124−12130.

Cocchi, M., Durante, C., Grandi, M., Lambertini, P., Manzini, D., & Marchetti, A. (2006). Simultaneous determination of sugars and organic acids in aged vinegars and chemometric data analysis. *Talanta, 69,* 1166-1175.

372 Cocchi, M., Lambertini, P., Mancini, D., Marchetti, A., & Ulrici, A. (2002). Determination of
373 carboxylic acids in vinegars and in aceto balsamico tradizionale di Modena by HPLC and GC
374 methods. *Journal of Agricultural and Food Chemistry, 50*, 5255-5261.
375 Consonni, R., & Cagliani, L.R. (2007). NMR relaxation data for quality characterization
376 of Balsamic vinegar of Modena. *Talanta, 73*, 332–339.
377 Consonni, R., Cagliani, L.R., Benevelli, F., Spraul, M., Humpfer, E., & Stocchero, M. (2008).
378 NMR and Chemometric methods: A powerful combination for characterization of Balsamic
379 and Traditional Balsamic Vinegar of Modena. *Analytica Chimica Acta*, *611*, 31–40.
380 Consonni, R., Cagliani, L.R., Rinaldini, S., & Incerti, A. (2008). Analytical method for
381 authentication of Traditional Balsamic Vinegar of Modena. *Talanta, 75*, 765–769.
382 Consonni, R., & Gatti, A. (2004). ^{1}H NMR Studies on Italian Balsamic and Traditional
383 Balsamic Vinegars. *Journal of Agricultural and Food Chemistry, 52*, 3446-3450.
384 Del Signore, A., Stancher, B., & Calabrese, M. (2000). Characterisation of balsamic vinegars
385 by amino acid content using a multivariate statistical approach. *Italian Journal of Food*
386 *Science, 12*, 317-332.
387 Dell'Oro, V., Ciambotti, A., & Tsolakis, C. (2012). Quality and safety of Italian wine vinegar.
388 *Italian Journal of Food Science, 24*, 182-187.
389 Gullo, C., Caggia, M., De Vero, L., & Giudici, P. (2006). Characterization of acetic acid
390 bacteria from traditional balsamic vinegar. *International Journal of Food Microbiology, 106*,
391 209-212.
392 Gullo, M., & De Vero, L. (2004). Ricerche finalizzate alla tutela della tipicità dell'Aceto
393 Balsamico Tradizionale di Reggio Emilia. (pp. 93–107). Reggio Emilia (Italia): Mucchi
394 Editore.
395 Harper, P.R., & Shahani, A.K. (2002). Modelling for the planning and management of bed
396 capacities in hospitals. *Journal of the Operational Research Society, 53*, 11–18.

Harper, P.R., & Winslett, D.J. (2004). Classification trees: a possible method for maternity risk grouping. *European Journal of Operational Research*, *169*, 146–156.

Henrion, R., & Henrion, G. (1994). Uberwachte klassifikation. In MultiVariate Datenanaly. (pp. 71–73). Berlin (Germany): Springer-Verlag.

Koda, M., Furihata, K., Wei, F., Miyakawa, T., & Tanokura, M. (2012). NMR-based metabolic profiling of rice wines by F-2-selective total correlation spectra. *Journal of Agricultural and Food Chemistry*, *60*, 4818−4825.

Krishnan, P., Kruger, N.J., & Ratcliffe, R.G. (2005). Metabolite fingerprinting and profiling in plants using NMR. *Journal of Experimental Botany*, *56*, 255-265.

Lachenmeier, D. W., Frank, W., Humpfer, E., Schäfer, H., Keller, S., Mörtter, M., & Spraul, M. (2005). Quality control of beer using high-resolution nuclear magnetic resonance spectroscopy and multivariate analysis. *European Food Research and Technology*, *220*, 215−221.

Loh, W.Y., & Vanichestakul, N. (1988). Tree-structured classification via generalized discriminant analysis (with discussion). *Journal of the American Statistical Association*, *83*, 715–728.

Lopez-Rituerto, E., Savorani, F., Avenoza, A., Busto, J. H., Peregrina, J. M., & Balling Engelsen, S. (2012). Investigations of La Rioja terroir for wine production using H-1 NMR metabolomics. *Journal of Agricultural and Food Chemistry*, *60*, 3452−3461.

Maes, P., Monakhova, Y. B., Kuballa, V., Reusch, H., & Lachenmeier, D. W. (2012). Qualitative and quantitative control of carbonated cola beverages using ^{1}H NMR spectroscopy. *Journal of Agricultural and Food Chemistry*, *60*, 2778−2784.

Mannina, L., D'Imperio, M., Gobbino, M., D'Amico, I., Casini, A., & Sobolev, A. P. (2012). Nuclear magnetic resonance study of flavoured olive oils. *Flavour and Fragrance Journal*, *27*, 250.

422 Masino, F., Chinnici, F., Franchini, G.C., Ulrici, A., & Antonelli, A. (2005). A study of the
423 relationships among acidity, sugar and furanic compound concentrations in set of casks for
424 Aceto Balsamico Tradizionale of Reggio Emilia by multivariate techniques *Food Chemistry*,
425 *92*, 673-679.
426 McLachlan, G.J. (1992). Discriminant Analysis and Statistical Pattern Recognition. New
427 York: John Wiley & Sons, Inc.
428 Papotti, G., Bertelli, D., Graziosi, R., Silvestri, M., Bertacchini, L., Durante, C., & Plessi, M.
429 (2012). Application of One- and Two-Dimensional NMR Spectroscopy for the
430 Characterization of Protected Designation of Origin Lambrusco Wines of Modena. *Journal of*
431 *Agricultural and Food Chemistry*, DOI: 10.1021/jf302728b.
432 Plessi, M., Bertelli, D., & Miglietta, F. (2006). Extraction and identification by GC-MS of
433 phenolic acids in traditional balsamic vinegar from Modena. *Journal of Food Composition*
434 *and Analysis*, *19*, 49-54.
435 Plessi, M., Monzani, A., & Coppini, D. (1989). Quantitative determination of acids and
436 derivates in balsamic and other vinegars. *Sciences des Aliments*, *9*, 179-183.
437 Ridley, S., Jones, S., Shahani, A., Brampton, W., Nielsen, M., & Rowan, K. (1998).
438 Classification trees for iso-resource grouping in intensive care. *Anaesthesia*, *53*, 833–840.
439 Satrioni F. (2010). Notiziario Della Consorteria Dell'aceto Balsamico Tradizionale
440 (http://www.museodelbalsamicotradizionale.org/ita/pages/consorteria/giornale-
441 settembre10.pdf) (17th October, 2012).
442 Savorani, F., Tomasi, G., & Engelsen, S. B. (2009). Icoshift: a versatile tool for the rapid
443 alignment of 1D NMR spectra. *Journal of Magnetic Resonance*, *202*, 190-202.
444 Theobald, A., Muller, A., & Anklam, E. (1998). Determination of 5-Hydroxymethylfurfural
445 in Vinegar Samples by HPLC. *Journal of Agricultural and Food Chemistry*, *46*, 1850-1854.

Thomas, F., & Jamin, E. (2009). ^{2}H NMR and ^{13}C-IRMS analyses of acetic acid from vinegar, ^{18}O-IRMS analysis of water in vinegar: International collaborative study report. *Analytica Chimica Acta*, *649*, 98-105.

Van-Diep, L., Zheng, X.-W., Ma, K., Chen, J.-Y., Han, B.-Z., & Nout, M. J. R. (2011). Characterization of Fen-Daqu through multivariate statistical analysis of ^{1}H NMR spectroscopic data. *Journal of the Institute of Brewing*, *117*, 516-522.

Winning, H., Rold - -Moreno, C., Cano, M. P., & Engelsen, S. B. (2009). An exploratory NMR nutri-metabonomic investigation reveals dimethyl sulfone as a dietary biomarker for onion intake. *Analyst*, *134*, 2344−2351.

Figure Captions

Figure 1. Typical ^{1}H NMR spectra recorded at 400 MHz of a TBVM sample; expansions of (A) aliphatic/alcoholic, and (B) sugar regions with metabolite assignments (see Table 2).

Figure 2. Comparison of the ^{1}H NMR spectra calibrated and normalized using the TMS signal of BVM (a), TBVM old (b) and TBVM extra old (c).

Figure 3. PCA score plot performed on complete spectral region data set: TBVM extra old (Δ), TBVM old (+), and BVM (□).

Figure 4. Vinegar classification using GDA on complete spectral region data set, showing the separation of samples in three evident clusters: TBVM extra old (Δ), TBVM old (+), BVM (□) and group centroids (●).

Figure 5. Vinegar classification using CTA (C&RT-style method) on the data set obtained by integrating the most significant signals resulted by GDA.

Table 1. Values of °Brix and *R ratio* in TBVM and °Brix in BVM samples

TBVM samples	Type[a]	°Brix	*R ratio*[b]	BVM samples	°Brix
1	Old	65	10.50	**41**	32
2	Old	67	12.91	**42**	22
3	Old	61.5	9.70	**43**	15.25
4	Old	69.5	11.94	**44**	26.5
5	Old	68.25	9.50	**45**	24
6	Extra old	71	10.88	**46**	19
7	Extra old	71.75	11	**47**	27
8	Extra old	71	11.14	**48**	23.5
9	Extra old	73	11.09	**49**	15.3
10	Extra old	70.5	12.56	**50**	20
11	Old	68.5	11.01	**51**	19.5
12	Extra old	73	9.92	**52**	19
13	Extra old	74	12.07	**53**	29.5
14	Old	70	11.59	**54**	38.5
15	Extra old	72	9.28	**55**	32
16	Extra old	71	11.75	**56**	36.5
17	Old	60.5	8.75	**57**	44.5
18	Old	65.5	9.86	**58**	40
19	Old	68	10.33	**59**	28.5
20	Extra old	72	12.16	**60**	39
21	Extra old	72.5	10.63	**61**	38.5
22	Extra old	70	10.08	**62**	53.5
23	Extra old	73	10.08	**63**	38.5
24	Extra old	72.5	10.54	**64**	42.5
25	Extra old	71.75	11.21	**65**	36
26	Old	65	9.88	**66**	18
27	Old	63	9.36	**67**	36
28	Old	64	10.24	**68**	31
29	Old	63.5	9.92	**69**	29
30	Old	65	10.25	**70**	30
31	Extra old	71.5	10.87	**71**	38.5
32	Extra old	71	9.89	**72**	53
33	Extra old	73.5	11.38	**73**	19
34	Extra old	72	11.41	**74**	38
35	Extra old	71	10.64	**75**	25
36	Extra old	72.5	11.49	**76**	30
37	Extra old	70.5	10.72		
38	Extra old	71	8.75		
39	Old	60.75	12.90		
40	Old	62	9.92		

[a]Age is indicated only for known aging process samples. Extra old >25 years; old >12 and <25 years. For BVM the aging is unknown, however it is < 3 years.
[b]°Brix/Total acidity % (p/v).

Table 2. Metabolites and ^{1}H chemical shifts identified[a]

Compound	**Group**	**δ (ppm)**	**Multiplicity[b]**	***J* (Hz)**
Acetic acid	$C2H_3$	1.90	s	-
Acetoin	$C4H_3$	1.13	d	7.0
	$C1H_3$	2.08	s	-
2,3-Butanediol	$C1H_3$	0.93	d	5.9
	$C4H_3$	0.98	d	5.9
Ethanol	$C2H_3$	1.04	t	7.2
Formic acid	HCOOH	8.16	s	-
β-Fructofuranose	C1H	3.44	m	-
	C6H	3.40	m	-
β-Fructopyranose	C1H	3.32; 3.48	dd	11.2; 5.4
	C3H C4H	3.62	m	-
	C5H	3.68	m	-
	C6H	3.82	d	12.0
6-Acetyl glucose signals[c]	CH_3CO	1.93-2.01	m	-
α–Glucopyranose	αC1H	4.89	d	3.5
	C4H	3.09	m	-
	C6H	3.62	m	-
β-Glucopyranose	βC1H	4.26	d	7.9
	C2H	2.87	t	8.4
5-HMF	C1H	9.56	s	-
	C3H	7.52	d	3.5
	C4H	6.63	d	3.3
	$C6H_2$	4.47	s	-
Lactic acid	$C3H_3$	1.19	d	6.7
Malic acid	C2H	2.38	m	-
	C2'H	2.59	m	-
	C3H	4.07	m	-
Succinic acid	$C2H_2$ $C3H_2$	2.40	s	-
Tartaric acid	C2H C3H	4.31	s	-
Valine	CγH3	0.87	d	6.8

[a]Assignments were from HSQC and HMBC experiments. The chemical shifts were expressed as relative values to those of TMS at 0 ppm.
[b]Peak multiplicities: s, singlet; d, doublet; t, triplet; dd, doublet of doublets; m, multiplet.
[c]Esters of glucose (6-acetylglucose) in the two anomeric forms. See Cirlini *et al.* for more details.[32]

Table 3. Calibration and validation results using CTA with building tree methods on the data set obtained by integrating the most significant signals resulted by GDA

Building tree method	**All samples calibration (n=76)**		**One-third validation (n=50+26)**					
			Calibration set		Validation set			
	Terminal nodes	*Correct classification (%)*	*Terminal nodes*	*Correct classification (%)*	*Correct classification (%)*			
					1	2	3	Total
DUS	23	100	12	100	75	62	100	81
DLCS	5	100	4	100	0	62	100	67
C&RT	12	100	9	100	100	50	100	81

Figure 1

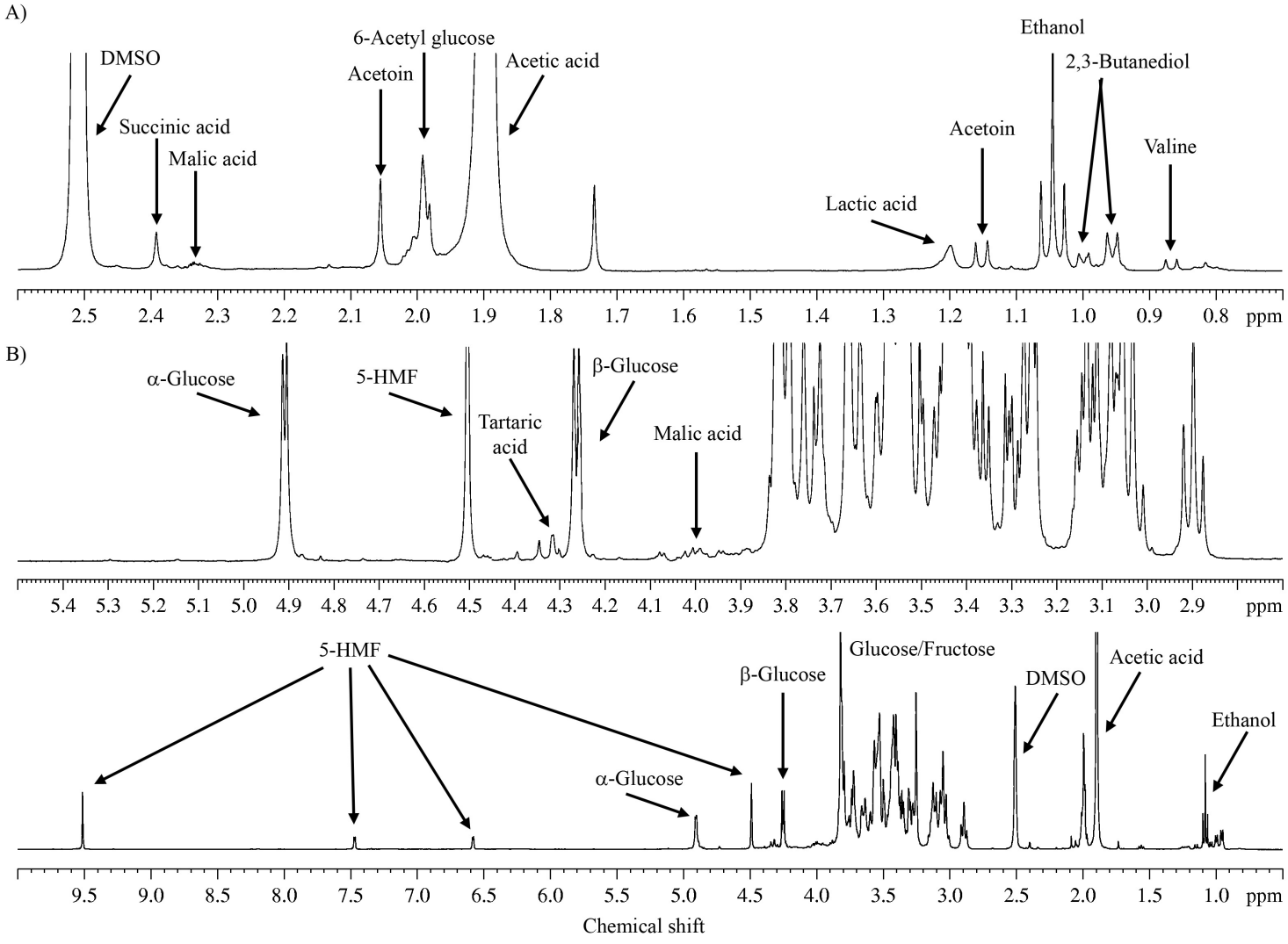

A)
DMSO
Succinic acid
Malic acid
6-Acetyl glucose
Acetoin
Acetic acid
Lactic acid
Acetoin
Ethanol
2,3-Butanediol
Valine
2.5 2.4 2.3 2.2 2.1 2.0 1.9 1.8 1.7 1.6 1.5 1.4 1.3 1.2 1.1 1.0 0.9 0.8 ppm
B)
α-Glucose
5-HMF
Tartaric acid
β-Glucose
Malic acid
5.4 5.3 5.2 5.1 5.0 4.9 4.8 4.7 4.6 4.5 4.4 4.3 4.2 4.1 4.0 3.9 3.8 3.7 3.6 3.5 3.4 3.3 3.2 3.1 3.0 2.9 ppm
5-HMF
α-Glucose
β-Glucose
Glucose/Fructose
DMSO
Acetic acid
Ethanol
9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 ppm
Chemical shift

Figure 2 color for web

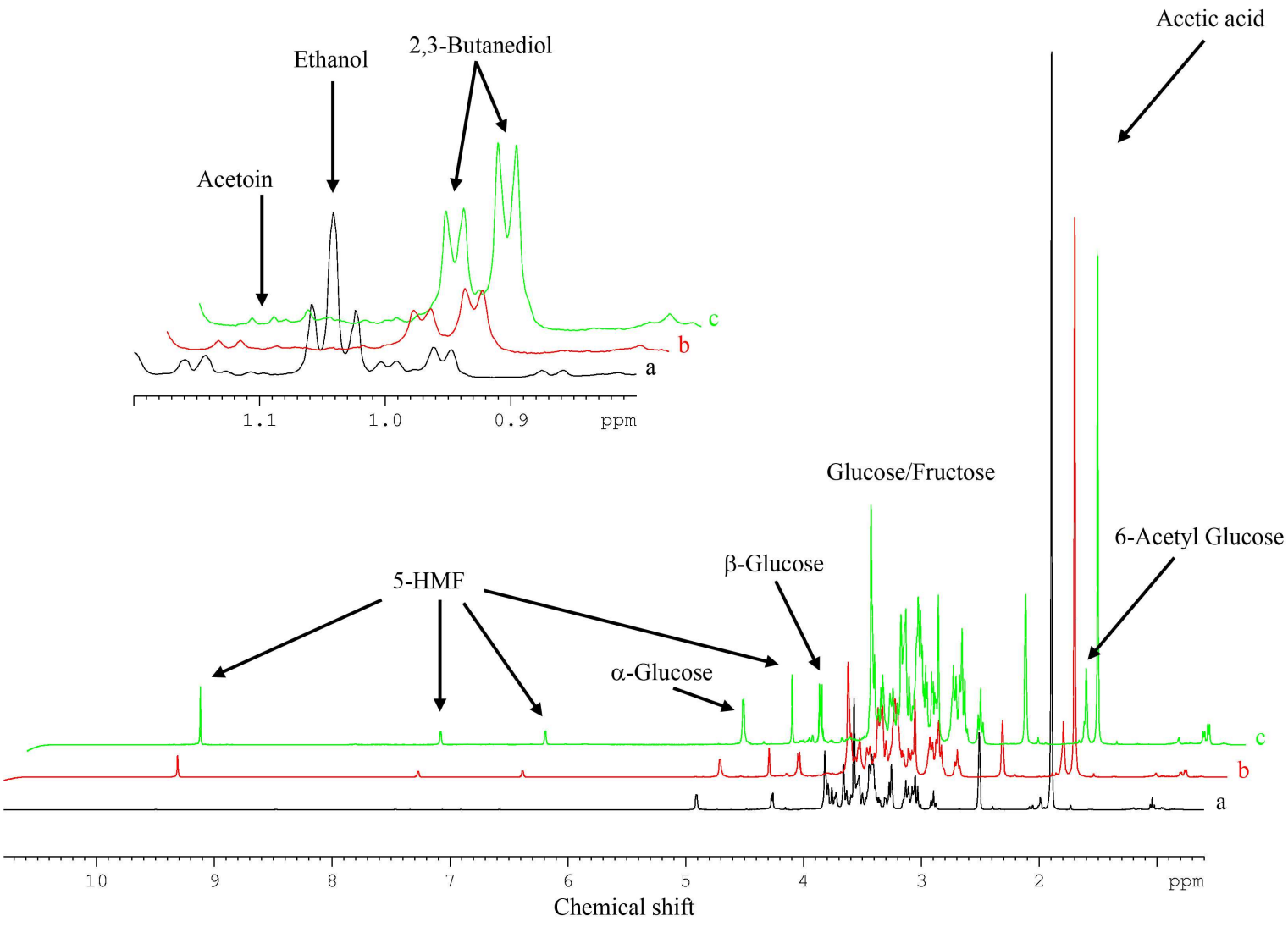


Figure 2 bw for print version

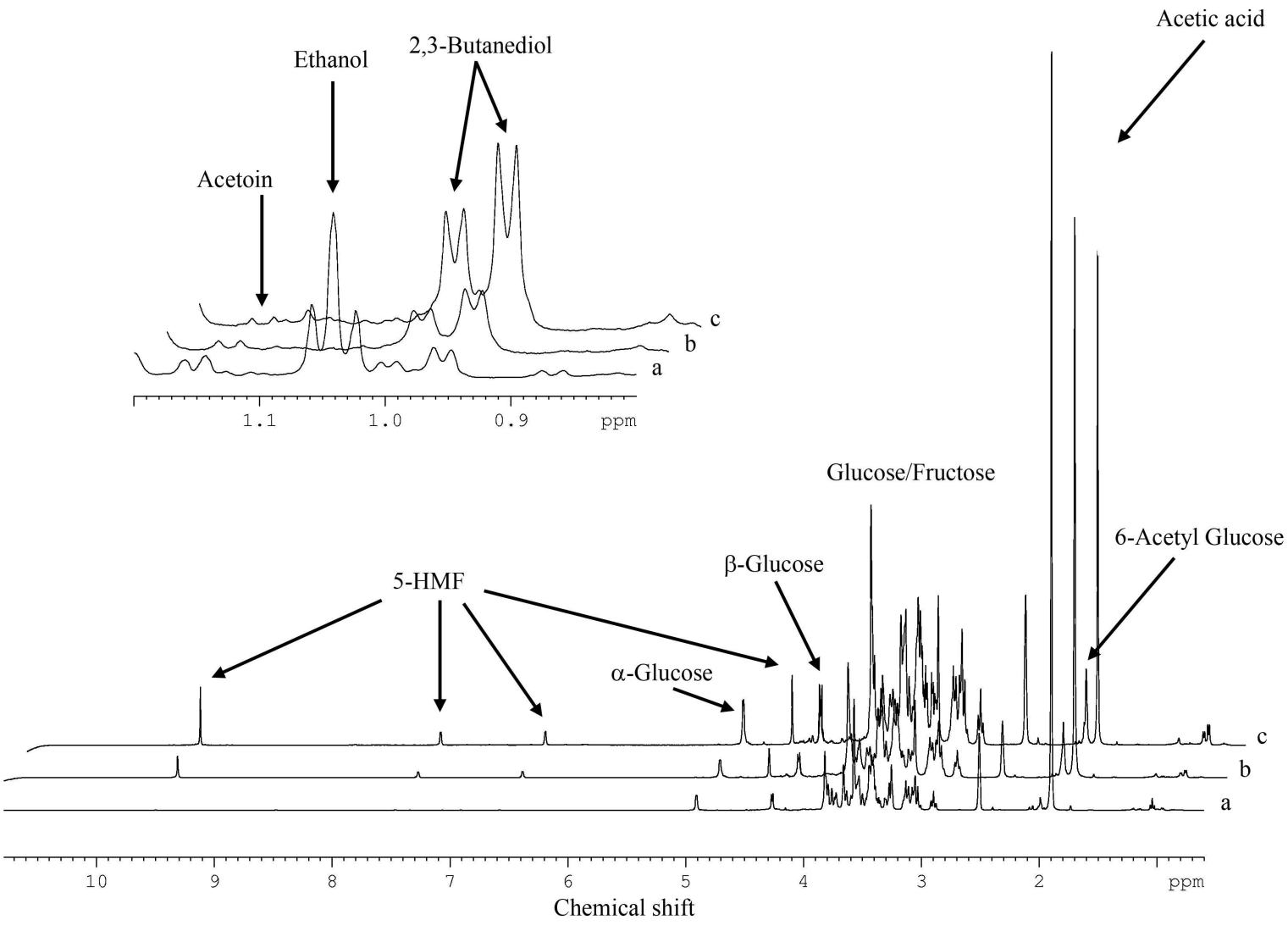


Figure 3

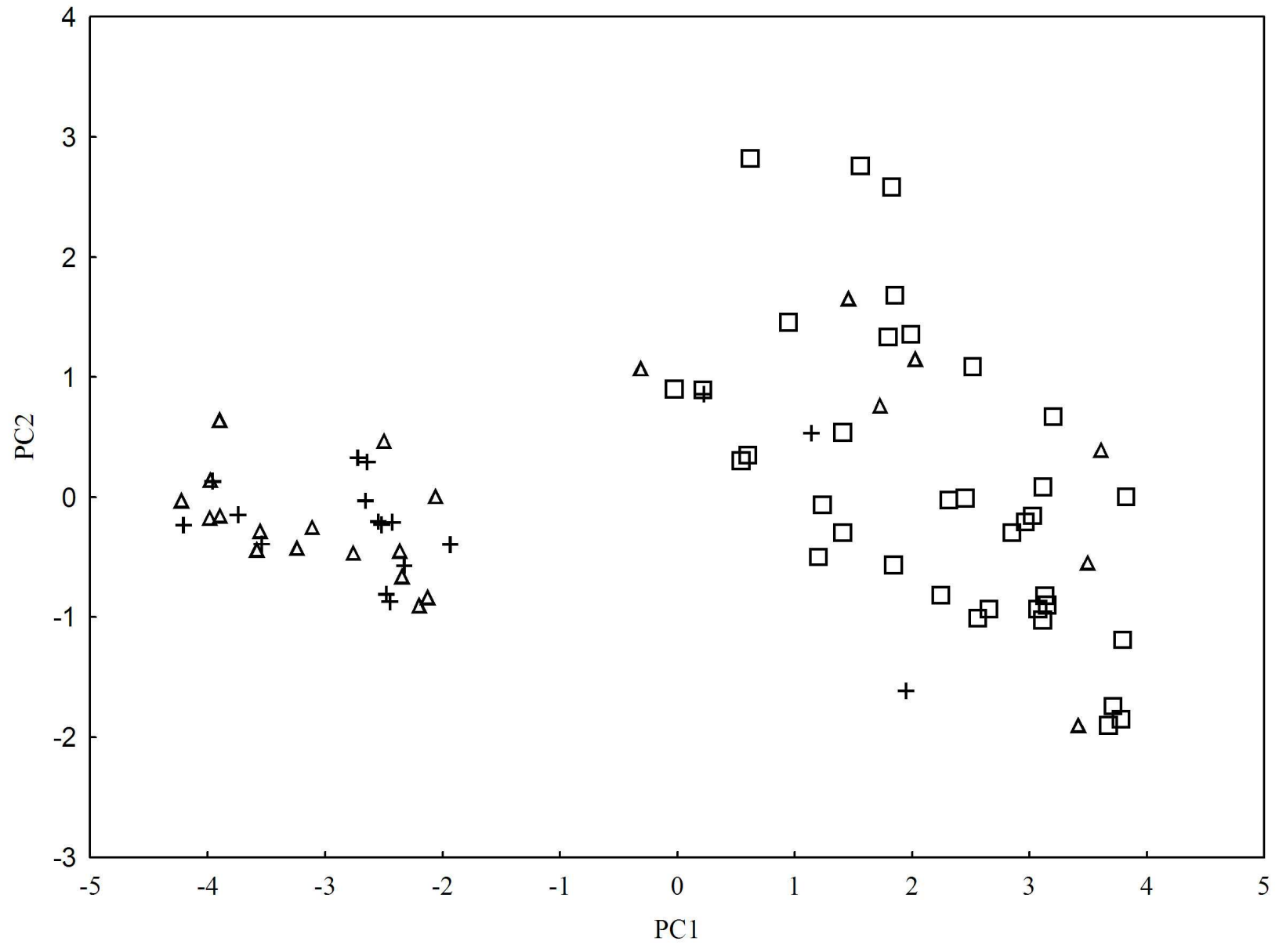


Figure 4

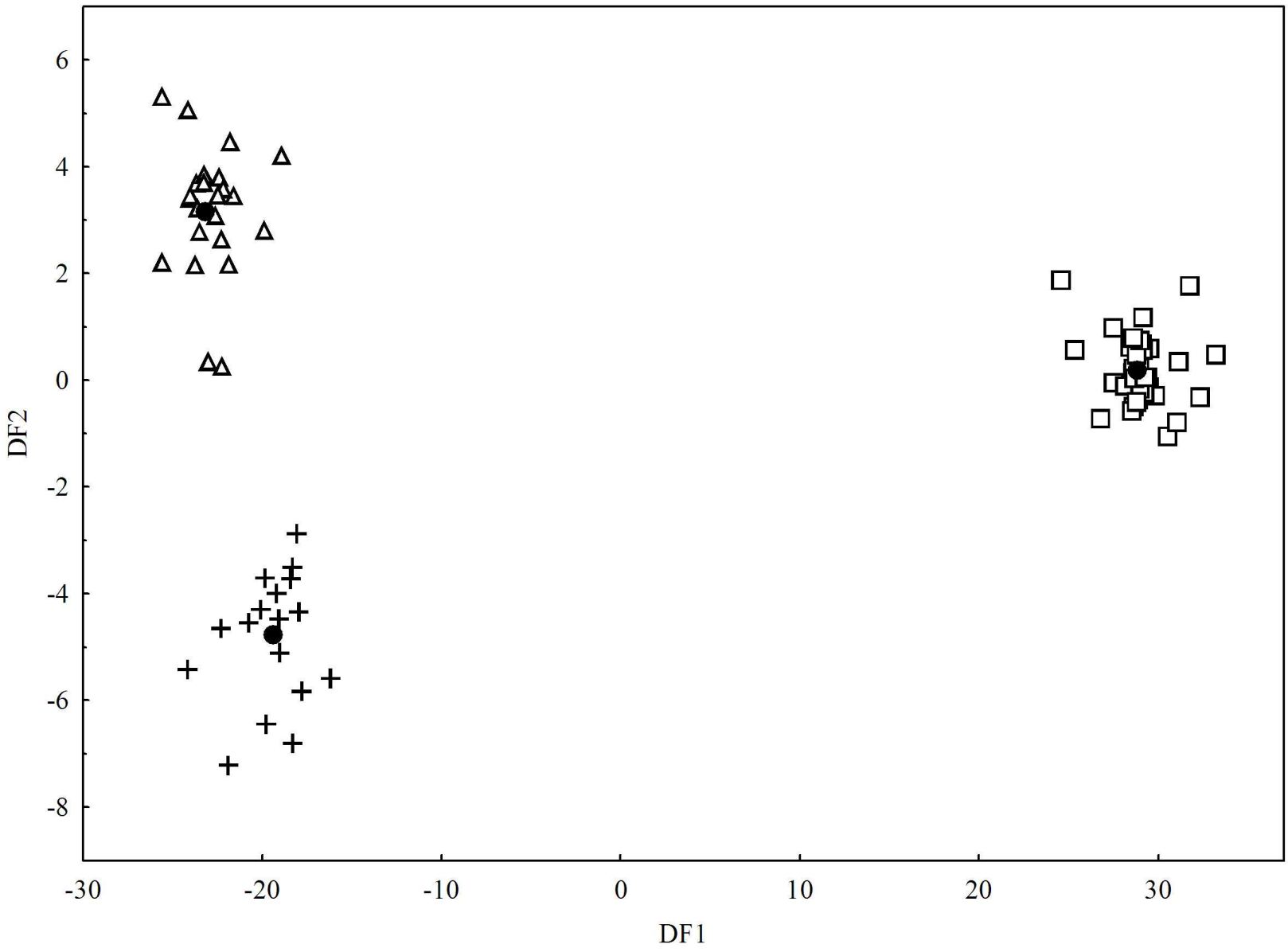


Figure 5 color for web

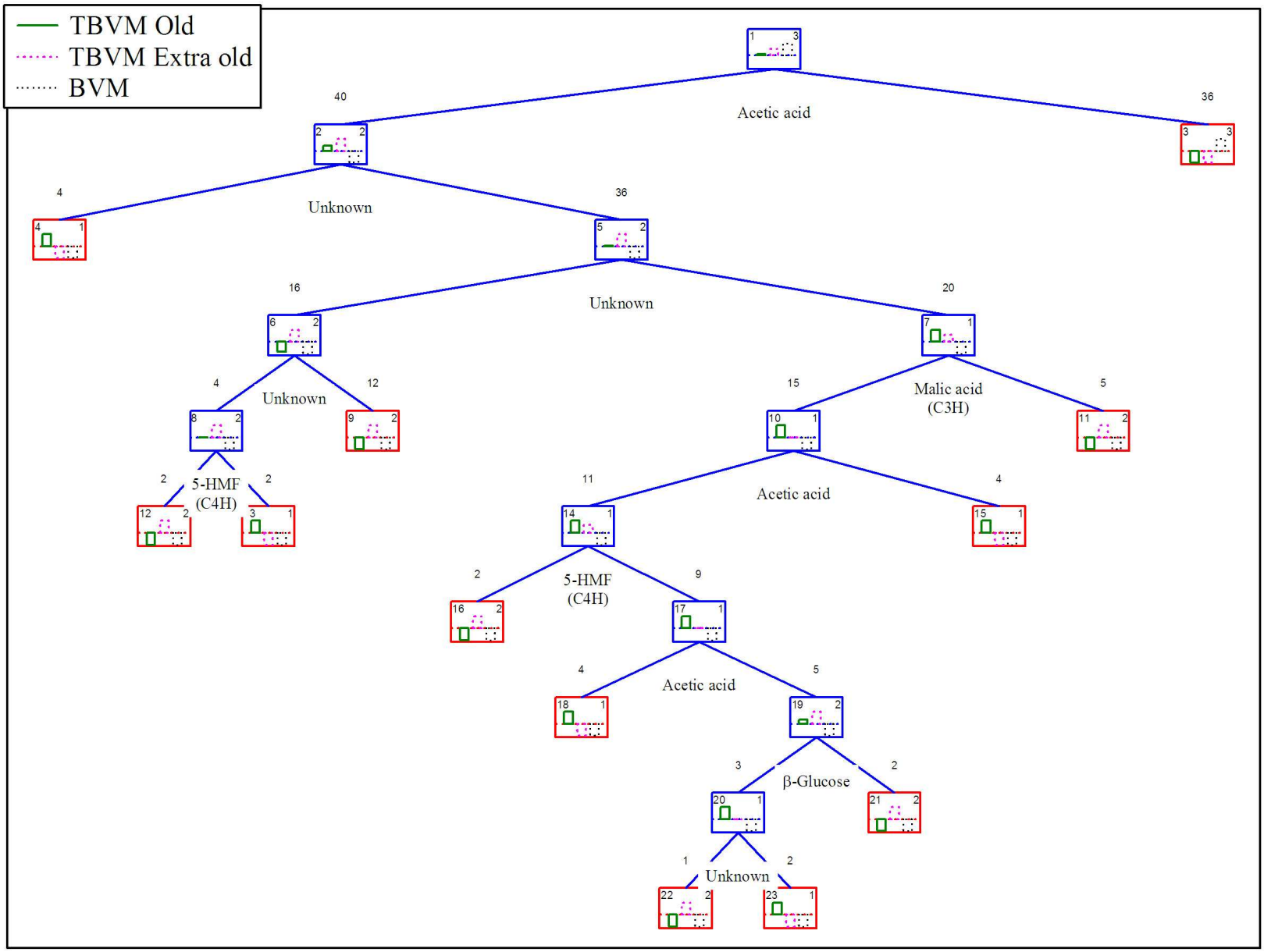
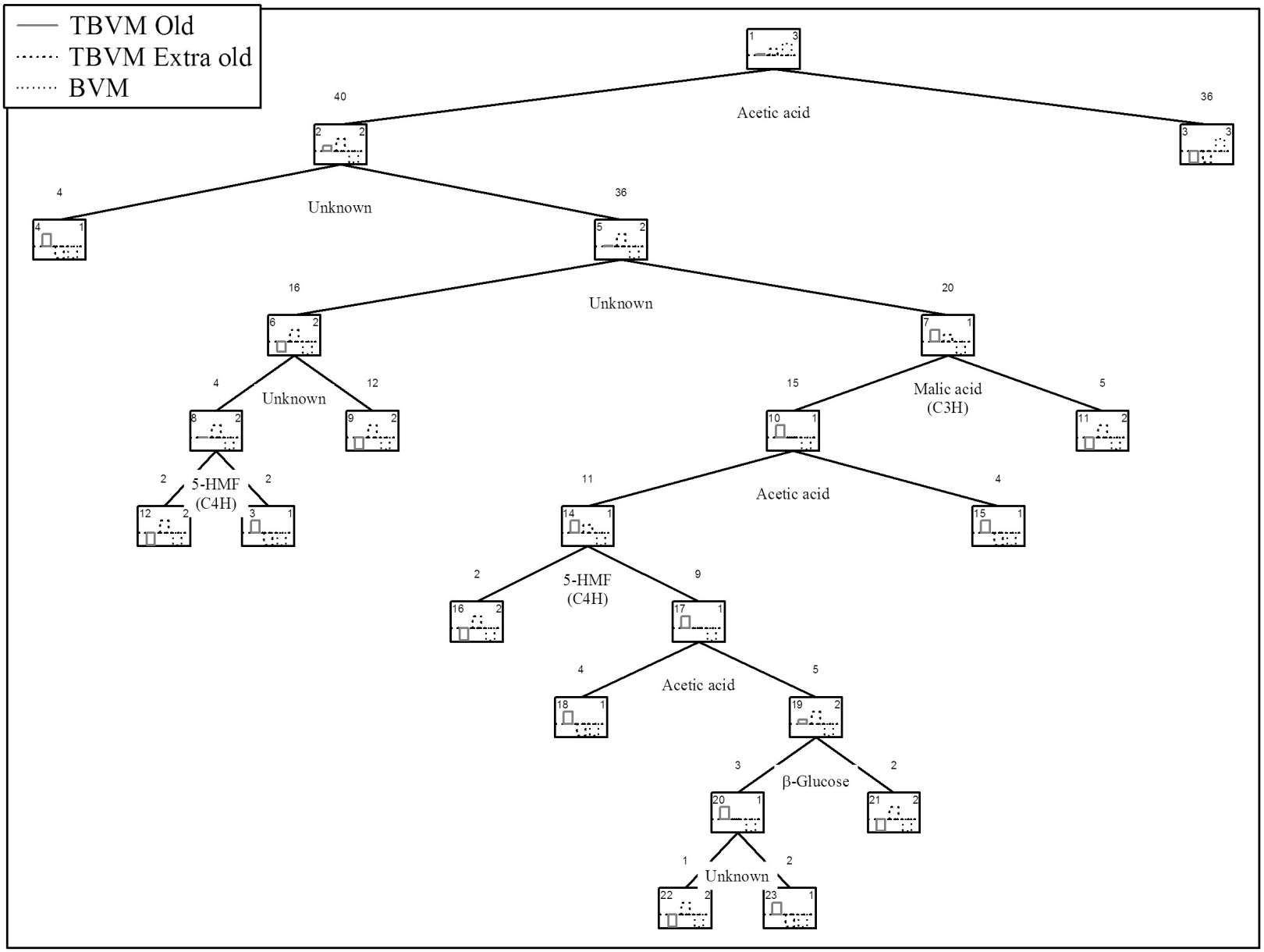


Figure 5 bw for print version



LWT-Food Science & technology

Manuscript LWT-D-13-00173

Traditional Balsamic vinegar and Balsamic vinegar of Modena analyzed by Nuclear Magnetic Resonance spectroscopy coupled with multivariate data analysis

Dear editor,

The manuscript was revised according to the comments

1. Number of keywords provided should be within 4-5.
   - Keywords are now 5;
2. Research papers should be within 20 pages of double-spaced type, (not including figures/tables and their captions).
   - The manuscript was revised and shortened to 19 pages;
3. Avoid vertical lines in tables.
   - The tables were corrected and vertical lines were deleted;