NMR spectroscopy and chemometrics-based analysis of grapevine

Ali, K.

CHAPTER 9

NMR metabolic fingerprinting-based identification of grapevine metabolites associated with Downy Mildew resistance

Kashif Ali[a], Federica Maltese[a], Eva Zyprian[b], Martina Rex[b], Young Hae Choi[a], and Robert Verpoorte[a]

[a] Natural Products Laboratory, Institute of Biology, Leiden University, The Netherlands
[b]Julius Kuehn Institute, Federal Research Centre for Cultivated Plants, Institute for Grapevine Breeding Geilweilerhof, Germany

Abstract

Grapevine (*Vitis vinifera ssp. vinifera* L.) and grapes have been extensively studied due to their numerous nutritional benefits and health affecting activities. In this study, metabolite fingerprinting of crude leaf extracts, based on ^{1}H nuclear magnetic resonance (NMR) spectroscopy and multivariate data analyses, has been used for the metabolic characterization of six different grapevine cultivars including downy and powdery mildew resistant `Regent´ and susceptible `Lemberger´ among others. Different two-dimensional (2D)-NMR techniques were also employed leading to the identification of different compounds. Principal component analysis (PCA), hierarchical cluster analysis (HCA), and projections to latent structures-discriminant analysis (PLS-DA) of the processed ^{1}H NMR data revealed clear differences among the cultivars. Metabolites responsible for the discrimination in different grapevine cultivars belong to organic acids, amino acids, carbohydrates, phenylpropanoids and flavonoids. A differentiation of the cultivars based on their resistance to downy mildew infection was also achieved and metabolites associated to this trait, namely, quercetin-3-*O*-glucoside and a *trans*-feruloyl derivative, were identified. On the basis of these results, the distribution of different plant metabolites among the different grapevine cultivars is presented.

Introduction

The Vitaceae family consists of almost one thousand species and among them *Vitis vinifera ssp. vinifera* is currently the most cultivated around the world (This et al. 2006). Grapevine and its products are also very well known as a source of bioactive compounds (Yilmaz and Toledo 2004), such as Vitamin E, flavonoids, stilbenoids, and procyanidins (also known as condensed tannins or oligomeric proanthocyanidins) (Iriti et al. 2006). The medicinal and nutritional value of grapes has been proclaimed for years which is supported by the identification of phytochemicals that exhibit a wide range of different activities such as antioxidant (Tedesco et al. 2000), anti-inflammatory/antiulcer (Saito et al. 1998), anticancer/antimutagenic (Nakagawa et al. 2001), antiobesity (Flechtner-Mors et al. 2004), apart from preventing cardiovascular diseases (Wallace et al. 2006), and dermal disorders (Khanna et al. 2001). Due to these vast uses and widespread cultivation, *Vitis sp.* has an enormous economical importance,

as a result of which all factors that affect its yield and quality as a crop are being intensely researched. Among these factors, their resistance to abiotic and biotic stress has been the subject of a great deal of studies (Christen et al. 2007; Oliveira et al. 2009). In their natural environment, plants are challenged by a number of potentially virulent microorganisms. The factors determining the resistance of plants against pathogens belong to some constitutive (structural barriers, phytoanticipins) and inducible defense mechanisms that include localized cell death, synthesis of phytoalexins, and pathogenesis-related proteins (Harborne 1999). Successful breeding programs of grape plants with increased resistance traits towards downy mildew should necessarily be based on a good understanding of the innate resistance mechanisms of cultivars against pathogenic fungi. In grapevine research, leaves and berries are the most targeted tissues for the pathology related studies, though leaves present the advantage of having a greater exposure to and thus interacting more with the pathogens as compared to berries, which are season dependent and not available throughout the year.

The term 'metabolome' has been used to describe the observable chemical profile or fingerprint of the metabolites present in whole tissues. Chemical analysis techniques applied to metabolites profiling should be unbiased, rapid, reproducible, and stable over time, while requiring only simple sample preparation. Many platforms are being used for the high throughput analysis of plant metabolites, but vary according to their sensitivity (Kopka et al. 2004). A technique that potentially meets all the above requirements is Nuclear Magnetic Resonance Spectroscopy (NMR). NMR has been widely used as a fingerprinting tool with multivariate or pattern recognition techniques such as the well-known principal components analysis (PCA) (Sumner et al. 2003). Recently, the combination of NMR and PCA has been applied to the metabolic profiling of various types of samples (Belton et al. 1998; Bresica et al. 2002; Charlton et al. 2002; Choi et al. 2004a; Duarte et al. 2002). This technique has proved to be a very powerful tool for the characterization of different species (Choi et al. 2005; Kim et al. 2005), and cultivars (Choi et al. 2004b). However, literature concerning metabolomic studies of the grapevine leaves using NMR is scarce. Recently, the method coupled with a transcriptional analysis, has been applied for the profiling of two grape cultivars with different resistance capabilities against pathogenic fungi (Figueiredo et al. 2008).

In this study, we identified the major metabolites contributing to the discrimination among six different grapevine cultivars using NMR spectroscopy and multivariate data analysis. Additionally, the metabolites which discriminate the cultivars on the basis of their capacity to resist downy mildew infection were also analyzed and their relative quantities were also determined.

Materials and Methods

Plant material and extraction

Leaves of different age from six cultivars of grapevine, grown in the green house of Julius Kuehn Institute, were used in this experiment (Table 1). Samples consisting of different leaves from the same plant of each cultivar were homogenized and analyzed in four replicates. The resistance of the grapevine cultivars towards *Plasmopara viticola* was measured using the nine-step classification of Organisation Internationale de la Vigne et du Vin (OIV, International Wine and Vine Organisation, http://www.oiv.int). A sample of 50 mg of freeze dried plant material was transferred to a 2 mL Eppendorf tube to which 1.5 mL of methanol-d_4:D_2O (KH_2PO_4 buffer, pH 6.0) (1:1) containing 0.05% TMSP (trimethyl silyl propionic acid sodium salt, w/v) was added. The mixture was vortexed at room temperature for 1 min, ultrasonicated for 20 min, and centrifuged at 13,000 rpm at room temperature for 5 min. An aliquot of 800 µL of the supernatant was transferred to a 5 mm NMR tube.

NMR measurements

^{1}H NMR and 2D *J*-resolved spectra were recorded at 25 °C on a 500 MHz Bruker DMX-500 spectrometer (Bruker, Karlsruhe, Germany) operating at a proton NMR frequency of 500.13 MHz. MeOH-d_4 was used as the internal lock. Each ^{1}H NMR spectrum consisted of 128 scans requiring 10 min and 26 s acquisition time with the following parameters: 0.16 Hz/point, pulse width (PW) = 30°, and relaxation delay (RD) = 1.5 s. A pre-saturation sequence was used to suppress the residual H_2O signal with low power selective irradiation at the H_2O frequency during the recycle delay. FIDs were Fourier transformed with LB = 0.3 Hz. The resulting spectra were manually phased and baseline corrected, and calibrated to TSP at 0.0 ppm, using XWIN NMR (version 3.5, Bruker). 2D *J*-resolved NMR spectra were acquired using 8 scans per 128

increments for F1 and 8 k for F2 using spectral widths of 5000 Hz in F2 (chemical shift axis) and 66 Hz in F1 (spin–spin coupling constant axis). A 1.5 s relaxation delay was employed, giving a total acquisition time of 56 min. Datasets were zero-filled to 512 points in F1 and both dimensions were multiplied by sine-bell functions (SSB = 0) prior to double complex FT. *J*-Resolved spectra tilted by 45°, were symmetrized about F1, and then calibrated, using XWIN NMR (version 3.5, Bruker). ^{1}H–^{1}H correlated spectroscopy (COSY) and heteronuclear multiple bonds coherence (HMBC) spectra were recorded on a 600 MHz Bruker DMX-600 spectrometer (Bruker). The COSY spectra were acquired with 1.0 s relaxation delay, 6361 Hz spectral width in both dimensions. Window function for COSY spectra was sine-bell (SSB = 0). The HSQC spectra were obtained with 1.0 s relaxation delay, 6361 Hz spectral width in F2 and 27,164 Hz in F1. Qsine (SSB = 2.0) was used for the window function of the HSQC. The HMBC spectra were recorded with the same parameters as the HSQC spectra except for 30,183 Hz of spectral width in F2. The optimized coupling constants for HSQC and HMBC were 145 Hz and 8 Hz, respectively.

Data analysis

The ^{1}H NMR spectra were automatically reduced to ASCII files. Spectral intensities were scaled to total intensity and reduced to integrated regions of equal width (0.04 ppm) corresponding to the region of δ 0.4–10.0. The regions of δ 4.75–4.9 and δ 3.28–3.34 were excluded from the analysis because of the residual signal of D_2O and CD_3OD, respectively. Bucketing was performed by AMIX software (Bruker). Principal component analysis (PCA) and projections to latent structures-discriminant analysis (PLS-DA) with scaling based on Pareto and Unit Variance method, respectively, were performed. Hierarchical cluster analysis (HCA) was also done using Ward's minimum variance method. All these analyses were performed with the SIMCA-P+ software (v. 12.0, Umetrics, Umeå, Sweden). The *t*-test for the ^{1}H-NMR signals was performed by MultiExperiment Viewer (v. 4.0) (Saeed et al. 2003).

Results

Metabolite identification

The establishment of a database of metabolites under well defined conditions, aimed at increasing the knowledge on the biological variability of a set of samples, is considered to be the key step of any metabolic study. A substantial part of this process is the identification of metabolites of different types, using appropriate methods. Several analytical tools are generally used in metabolomics but MS and NMR based studies are the most widely accepted. Though not highly sensitive, ^{1}H NMR, is increasingly chosen now because of the simple sample preparation required and its possibility of detecting very dissimilar groups of metabolites in fairly short periods of time. Both 1D and 2D NMR techniques allow the identification of different classes of compounds including amino acids, carbohydrates, organic acids, and phenolic compounds thus facilitating the recognition of a wide-ranging metabolome.

Although the use of NMR in metabolomic studies has many advantages, the overlapping of the signals in NMR spectra represents a major difficulty in the identification of each metabolite. The problem is usually overcome by obtaining 2D ^{1}H–^{1}H *J*-resolved NMR spectra to provide additional information of each signal. The use of *J*-resolved spectra together with other correlation 2D-NMR spectra such as COSY, HSQC, and HMBC significantly increased the number of identified metabolites in this study. Moreover, results were confirmed by comparison with NMR spectra of the corresponding reference compounds. This therefore afforded an evaluation of the variation in the content of these compounds throughout the six different cultivars analyzed by NMR (Table 1), reflecting the metabolome of each sample.

This study allowed the identification of a number of different metabolites in the leaves of different grapevine cultivars using ^{1}H NMR and 2D *J*-resolved spectra together with ^{1}H-^{1}H COSY, and HMBC spectra. These metabolites included amino acids, organic acids, carbohydrates, flavonoids, and phenylpropanoids (Table 2, Figure 1). The ^{1}H NMR spectra can be divided into three distinct regions, one of which, the area between δ 0.8-4.0, corresponds to amino acids and organic acids. The high signal intensity in this region helped elucidate the signals by comparison with reference spectra. This region showed signals of methionine, proline, valine, threonine, leucine, alanine, glutamine, glutamic acid, α-linolenic acid, acetic acid and succinic acid. The region between δ 4.0-5.5 is considered to be the carbohydrate region and in this case the signal of the anomeric protons of β-glucose, α-glucose, fructose, and sucrose were detected.

Table 1. Variety names, their codes, and characteristics, used in this experiment.

S. No.	Working Code	Variety Name	Leaf OIV 452	Characteristics
1	RG1	'Regent'	8[a]	Resistant
2	RG2	Gf. Ga. 47-42	6[a]	Resistant
3	RG3	'Villard blanc'	8[a]	Resistant
4	RG4	'Boerner'	8[a]	Resistant
5	SG5	'Lemberger'	~	Susceptible
6	SG6	V3125	~	Susceptible

[a]Average data over a period of 12 years (from JKI-Institute for Grapevine Breeding, pers. communication R. Eibach).
~ Variety based on *V. vinifera* which is susceptible against *Plasmopara viticola*.

In the aromatic region, the low signal intensity and lack of reference compounds spectra were the main obstacles for compound identification. Signal overlapping also caused difficulties but this problem was overcome employing 2D NMR techniques. The aromatic region showed the presence of major doublets (16.0 Hz) in the range of δ 6.4–6.5 and also in the region of δ 7.6-7.7, which are typical signals of H-8' and H-7' of phenylpropanoids, respectively (Figure 2A). The COSY spectra also confirmed the correlation between H-8' and H-7' of phenylpropanoids (Figure 2B), with the coupling with carbonyl carbon at δ 171 in the HMBC spectra. A *trans*-phenylpropanoid and its *cis* form were elucidated by two dimensional NMR i.e. *trans*-feruloyl derivative at δ 6.47 (d, J = 16.0 Hz), δ 6.87 (d, J = 8.4 Hz), δ 7.06 (dd, J = 8.4, 2.3 Hz), δ 7.26 (d, J = 2.0 Hz), δ 7.65 (d, J = 16.0 Hz) and *cis*-feruloyl derivative at δ 5.97 (d, J = 13.0 Hz), δ 6.84 (d, J = 8.8 Hz), δ 6.94 (d, J = 13.0 Hz), δ 7.13 (dd, J = 8.4, 2.0 Hz), δ 7.83 (d, J = 2.0 Hz).

However, the *cis*- forms of phenylpropanoids are considered as to be artifacts of their *trans*- forms possibly produced during sample extraction or storage (Liang et al. 2006). Additionally, a compound formed between tartaric acid and a feruloyl moiety, some form of feruloyl tartaric acid was detected, but the very low signal intensity hindered the complete assignments for this compound.

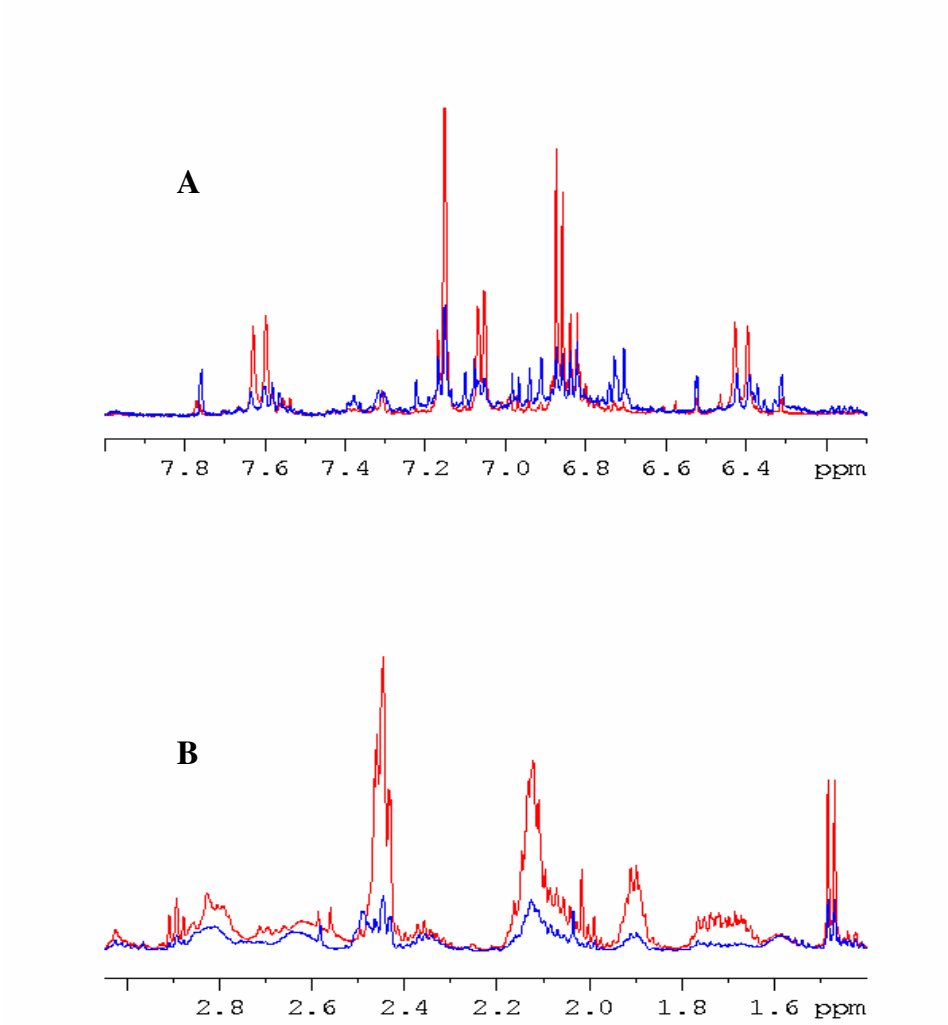

Figure 1. ^{1}H NMR spectra of two varieties of *Vitis* species showing phenolic (A) and amino acid (B) regions. The resistant cultivar shows higher phenolic and amino acid contents than susceptible cultivar.

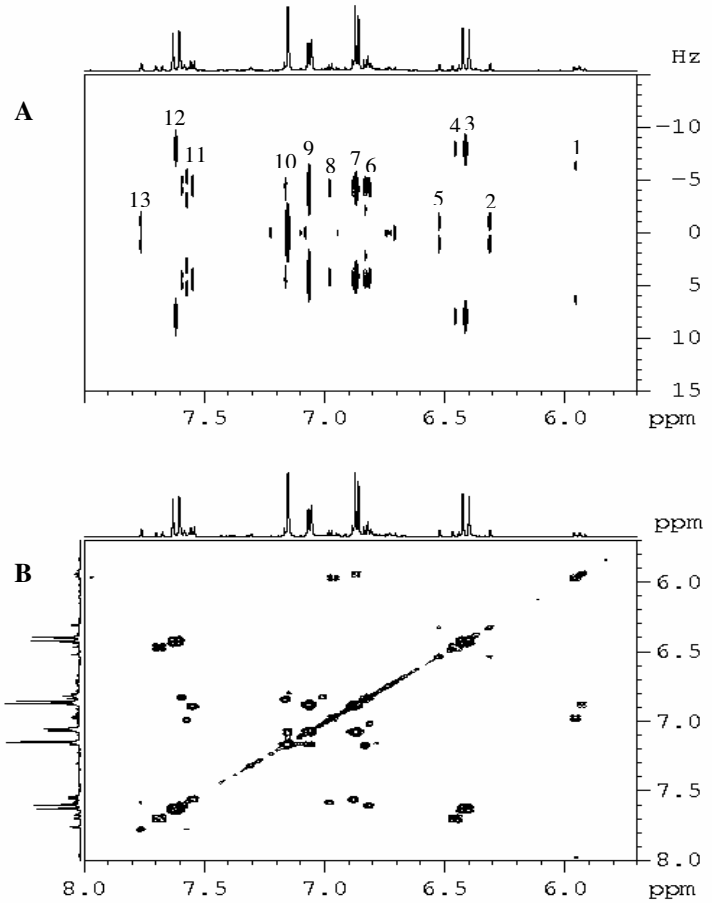

Figure 2. Two dimensional *J*-resolved (A) and ^{1}H-^{1}H COSY (B) spectra of grapevine leaf extracts in the region of d 8.0 - 5.7. In *J*-resolved spectrum (A) the following signals are observed. 1, H-8': *cis*-feruloyl derivative; 2, H-8: Quercetin-3-*O*-glucoside; 3, H-8': Caffeic acid; 4, H-8': *trans*-feruloyl derivative; 5, H-6: Quercetin glucoside; 6, H-5': Caffeic acid; 7, H-3: *trans*-feruloyl derivative; 8, H-5': Quercetin glucoside; 9, H-6': Caffeic acid; 10, H-6': *cis*-feruloyl derivative; 11, H-6': Quercetin glucoside; 12, H-7: Caffeic acid; 13, H-2': Quercetin glucoside. In COSY spectrum (B) the correlations of H-7' (δ 7.65) and H-8' (δ 6.47) of *trans*-feruloyl derivative; H-5' (δ 6.97) and H-6' (δ 7.56), and H-6 (δ 6.53) and H-8 (δ 6.32) of quercetin-3-*O*-glucoside are observed.

Another phenylpropanoid was identified using 1D and 2D NMR Spectra. Signals at δ 7.60 (d, J = 2.0 Hz), δ 7.12 (d, J = 2.0 Hz), δ 7.06 (dd, J = 8.0, 2.0 Hz), δ 6.86 (d, J = 8.0 Hz), δ 6.40 (d, J = 16.0 Hz) in ^{1}H NMR, are associated to a caffeoyl moiety. In this particular caffeoyl moiety, correlation between signals of H-9 at δ 7.60 and H-10 at δ

6.40 was observed in the COSY spectrum. The COSY spectrum also showed the correlation of signals at δ 7.12 with δ 6.80 and δ 7.06 for H-6, H-4, and H-3, respectively. The attachment of tartaric acid to the caffeoyl moiety was confirmed by a downfield shift of the typical tartaric acid signal from δ 4.80 to δ 5.34 due to its bonding to the carboxylic function of caffeic acid, which was also correlated with the signal at δ 168.4 in HMBC spectrum. A $-OCH_3$ signal at δ 3.71 also correlated with the tartaric acid signal at δ 5.34 which implies the attachment of a methyl group to the tartaric acid. Based on these assignments the compound was identified as caffeoyl tartaric acid methyl ester.

Compounds such as fumaric acid at δ 6.60 (s), tartaric acid at δ 4.35 (s), and shikimic acid at δ 6.56 (dt, J = 4.0, 1.7 Hz), δ 4.32 (t, J = 4.5 Hz), δ 4.00 (m), δ 3.70 (dd, J = 10.0, 4.0 Hz), δ 2.69 (dt, J = 18.0, 5.0 Hz), δ 2.18 (dt, J = 18.1, 1.7 Hz) were also recognized in the phenolic region. Different protons of shikimic acid, i.e, H-4 (δ 6.56) with H-3 (δ 4.32) and H-3 with H-2 (δ 3.70) were coupled. The signals at δ 2.69 and δ 2.18 were correlated, in the COSY spectrum, not only with each other but also with a signal at δ 4.00 (H-1) of shikimic acid.

A flavanoid, quercetin-3-*O*-glucoside, with signals at δ 7.77 (d, J = 2.0 Hz), δ 7.56 (dd, J = 8.0, 2.0 Hz), δ 6.97 (d, J = 8.5 Hz), δ 6.53 (d, J = 2.0 Hz), δ 6.32 (d, J = 2.0 Hz) was also identified in the aromatic region. As shown in Figure 2B, the signal at δ 6.53 of H-6 (d, J = 2.0 Hz) was correlated in the COSY spectrum with the signal at δ 6.32 of H-8 (d, J = 2.0 Hz) and a signal at δ 6.97 of H-5′ (d, J = 8.8 Hz) with one at δ 7.56 of H-6′ (dd, J = 8.0, 2.0 Hz). All of these assignments (Table 2) were done by comparing the spectra with 1D and 2D NMR spectra of common plant metabolites in our in-house library.

Multivariate data analyses

Principal component analysis (PCA) is an unsupervised, unbiased, and clustering method used to reduce the dimensionality of multivariate data. The principal components (PCs) can be exhibited in a graphical form as a "score plot". This plot is useful for the identification of any groupings in the data set and is also used to highlight outliers that may be due to errors in sample preparation or instrumentation parameters. Coefficients by which the original variables must be multiplied to obtain the PCs are

called "loadings" (Goodacre et al. 2000). Thus for NMR data, "loading plots" can be used to detect the spectral areas responsible for the separation in the data.

Table 2. ^{1}H NMR chemical shifts (δ) and coupling constants (Hz) of grapevine metabolites.

Compounds	**Chemical Shifts (δ) and Coupling Constants (Hz)**
Methionine	2.14 (s), 3.79 (t, *J*=6.0)
Proline	2.35 (m), 3.37 (m)
Valine	1.01 (d, *J*=7.0), 1.06 (d, *J*=7.0), 2.28 (m)
Leucine	0.98 (d, *J*=7.5), 0.96 (d, *J*=7.5)
Threonine	1.32 (d, *J*=7.0), 3.51 (d, *J*=12.0), 4.27 (m)
Alanine	1.48 (d, *J*=7.4), 3.73 (q, *J*=7.4)
Glutamine	2.14 (m), 2.41 (td, *J*=16.2, 7.5)
Glutamic acid	2.13 (m), 2.42 (m), 3.71 (dd, *J*=7.0, 1.9)
Inositol	4.01 (t, *J*=2.8)
Sucrose	4.16 (d, *J*=8.1), 5.39 (d, *J*=3.9)
α-Glucose	5.17 (d, *J*=3.8)
β-Glucose	4.58 (d, *J*=7.9)
Rhamnosyl moiety	1.10 (d, *J*=6.1)
Adenine	8.19 (s), 8.22 (s)
Fumaric acid	6.60 (s)
Tartaric acid (free)	4.35 (s)
Gallic acid	7.04 (s)
Succinic acid	2.53 (s)
Ascorbic acid	4.52 (d, *J*=2.0)
α-Linolenic acid	0.95 (t, *J*=7.5)
Acetic acid	1.94 (s)
Caffeoyl tartaric acid methyl ester	6.40 (d, *J*=16.0), 6.86 (d, *J*=8.0), 7.06 (dd, *J*=8.0, 2.0), 7.12 (d, *J*=2.0), 7.60 (d, *J*=16.0), 5.46 (d, *J*=2.6), 3.71 (s)
Shikimic acid	2.18 (dt, *J*=18.1, 1.7), 2.69 (dt, *J*=18, 5), 3.70 (dd, *J*=10, 4), 4.00 (m), 4.32 (t, *J*=4.5), 6.56 (dt, *J* = 4.0, 1.7 Hz)
1-*O*-ethyl-β-glucoside	1.19 (t, *J* = 7.0)
Myricetin	6.30 (d, *J*=2.0), 6.52 (d, *J*=2.0), 7.3 (s)
Quercetin-3-*O*-glucoside	6.32 (d, *J*=2.0), 6.53 (d, *J*=2.0), 6.97 (d, *J*=8.5), 7.56 (dd, *J*=8.0, 2.0), 7.77 (d, *J*=2.0), 5.30 (d, *J*=7.6)
cis-Feruloyl derivative	5.97 (d, *J*=13.0), 6.84 (d, *J*=8.8), 6.94 (d, *J*=13.0), 7.13 (dd, *J*=8.4, 2.0), 7.83 (d, *J*=2.0)
trans-Feruloyl derivative	6.47 (d, *J*=16.0), 6.87 (d, *J*=8.4), 7.06 (dd, *J*=8.4, 2.3), 7.26 (d, *J*=2.0), 7.65 (d, *J*=16.0)

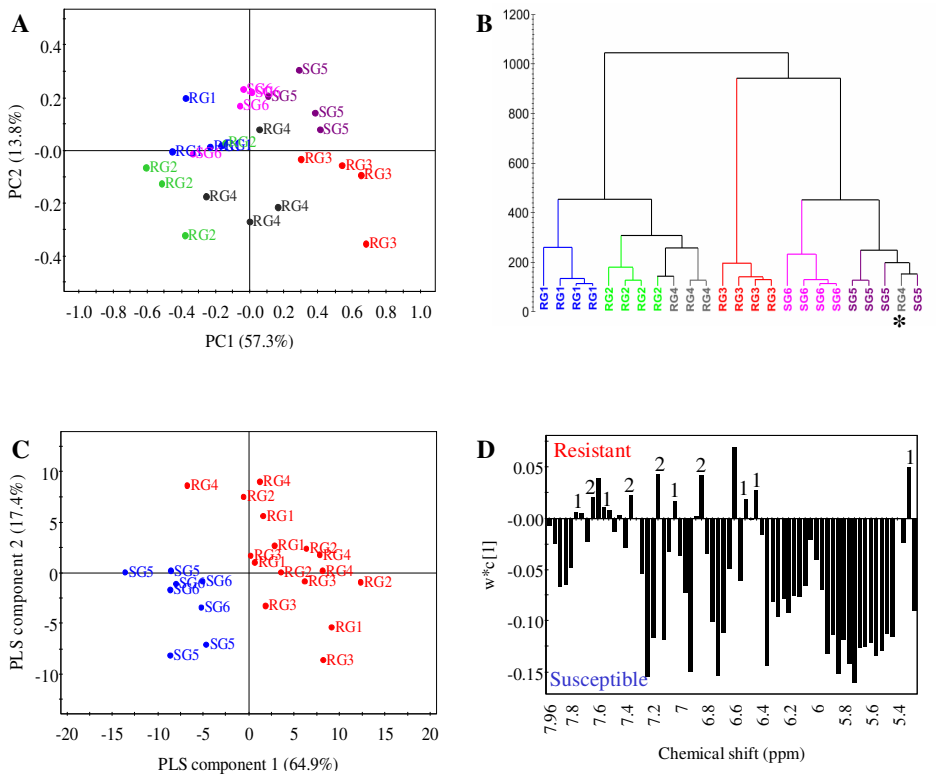

Figure 3. Score plot of PCA (A), dendrogram of HCA using Ward's minimum variance Method (B), Score plot of PLS-DA (C), loading plot of PLS-DA (D), based on whole range of ^{1}H NMR signals (δ 0.3-10.0). Score plot (A) shows the compounds responsible for the separation of six grapevine cultivars. In HCA clustering (B) samples are clustered on the basis of similarity, the sample with * is an outlier. PLS-DA (C) shows resistant (red) and susceptible (blue) classes are separated. The loading plot of PLS-DA (D) shows the signals of compounds responsible for the separation of resistant and susceptible classes. 1: ^{1}H NMR signals of Quercetin-3-*O*-glucoside including δ 5.30 (d, *J*=7.6), δ 6.32 (d, *J*=2.0), δ 6.53 (d, *J*=2.0), δ 6.97 (d, *J*=8.5), δ 7.56 (dd, *J*=8.0, 2.0), δ 7.77 (d, *J*=2.0), 2: 1H NMR signals of *trans*-feruloyl derivative δ 6.47 (d, *J*=16.0), δ 6.87 (d, *J*=8.4), δ 7.06 (dd, *J*=8.4, 2.3), δ 7.26 (d, *J*=2.0), δ 7.65 (d, *J*=16.0).

The first part of this study consisted in the application of PCA in order to observe metabolic characters of the six grapevine cultivars. In the PCA score plot, all six cultivars were separated from each other by component 1 and 2 (Figure 3A). The SG5, SG6, and RG1 cultivars showed positive PC2 values but they differed in their PC1 values, being positive and negative, respectively. Cultivars like RG1, RG2, and RG4 showed one outlier each. The three resistant cultivars (RG2, RG3, and RG4) showed negative PC2 values and differed in their PC1 values as RG3 and 4 showed positive PC1 values while RG2 had negative PC1 values.

The separation observed in PCA can be explained in terms of the identified compounds, using the loading plots for PC1 and 2. Signals of caffeic acid and fructose were found to be discriminating for the variety SG6, while SG5 showed higher signals for linolenic acid and adenine. Quercetin-3-*O*-glucoside is proved discriminating for the RG3 variety. Three varieties, RG1, RG2 and RG4, were grouped closer indicating that they might possibly share their metabolic profile. Metabolites responsible for their separation were identified as amino acids such as alanine and proline, along with succinic acid and inositol. It is interesting that although many compounds were identified in the phenolic region, only few turned out to be responsible for the discrimination of these grapevine varieties.

Similar to PCA, Hierarchical cluster analysis (HCA) is an unsupervised method. In HCA, based on samples similarity or distance, progressive pair-wise grouping of samples is ocurred. Several distance measures like Euclidean distance, Manhattan distance, or correlation, can be used in HCA but the results of different measures will be accordingly different. The HCA results can be seen as a dendrogram in which branch lengths reflect the differences among the groups and thus provide an easy visualization of the similarities of samples (Sumner et al. 2003). Ward's method uses an analysis of variance approach to evaluate the distance between clusters. In general, this method is regarded as very efficient because it tends to create equally sized small clusters (Ward 1968). The results of the dendogram of HCA of the spectral data of the cultivars (Figure 3B) are quite similar to those obtained with PCA. Three resistant varieties, RG1, RG2 and RG4, were clustered together showing relative similarities and also that they share their metabolic profile. The remaining resistant variety, RG3, was relatively distant from both the susceptible as well as the other resistant varieties. This may be due to the high levels of quercetin-3-*O*-glucoside, as observed by the loading plot of PCA. As expected, the two susceptible varieties also show relative similarities and were grouped together. The height of the clusters are proportional to the distance (difference) between the clusters. That is, when the vertical lines are tall the clusters are far apart (different), and when they are short the clusters are close together (similar).

Both PCA score plot and HCA dendogram showed some outliers (Figure 3A and 3B). The possible reasons for the outliers in the multivariate data analyses is the production of artifacts due to sample storage, also the extraction solvents may have caused the

production of such artifacts during extraction (Verpoorte et al. 2008). Another very important reason is the age of the plant sample as the young and old leaves of the same plant can be different in their metabolic profile (Abdel-Farid et al. 2009).

The next step in the metabolomic study consisted in applying projections to latent structures-discriminant analysis (PLS-DA) which, unlike the unbiased system used for PCA, was performed on pre-input information. The most important information obtained from PLS-DA is the correlation between two data sets and in this case, the ^{1}H NMR signals and their classification as resistant and susceptible cultivars were investigated. The PLS-DA was applied by classifying the varieties into two groups. The first group was for the resistant varieties RG1, RG2, RG3, and RG4, and the second group was for the susceptible varieties, SG5 and SG6.

The PLS-DA completely separated both the resistant and susceptible groups by component 1 (Figure 3C). The metabolites responsible for that separation were identified as acetic acid, inositol, fumaric acid, and succinic acid for the resistant group. The metabolites adenine, fructose, glucose, ascorbic acid, caffeic acid, and shikimic acid were responsible for the separation of the susceptible varieties. In the case of resistant varieties, signals of two phenolic compounds, quercetin-3-*O*-glucoside and a *trans*-feruloyl derivative, were also found responsible for the separation as shown by the column plot (Figure 3D).

Metabolite quantification

The entire data set was submitted to the *t*-test for the confirmation and relative quantification of the signals responsible for separation in PLS-DA. The *t*-test confirmed that those metabolites discriminating the group of resistant and susceptible cultivars were indeed statistically significant ($p<0.05$). Figure 4 shows the relative quantity of these compounds in all six varieties. These quantities were measured on the basis of the mean peak areas of the characteristic signals of these compounds.

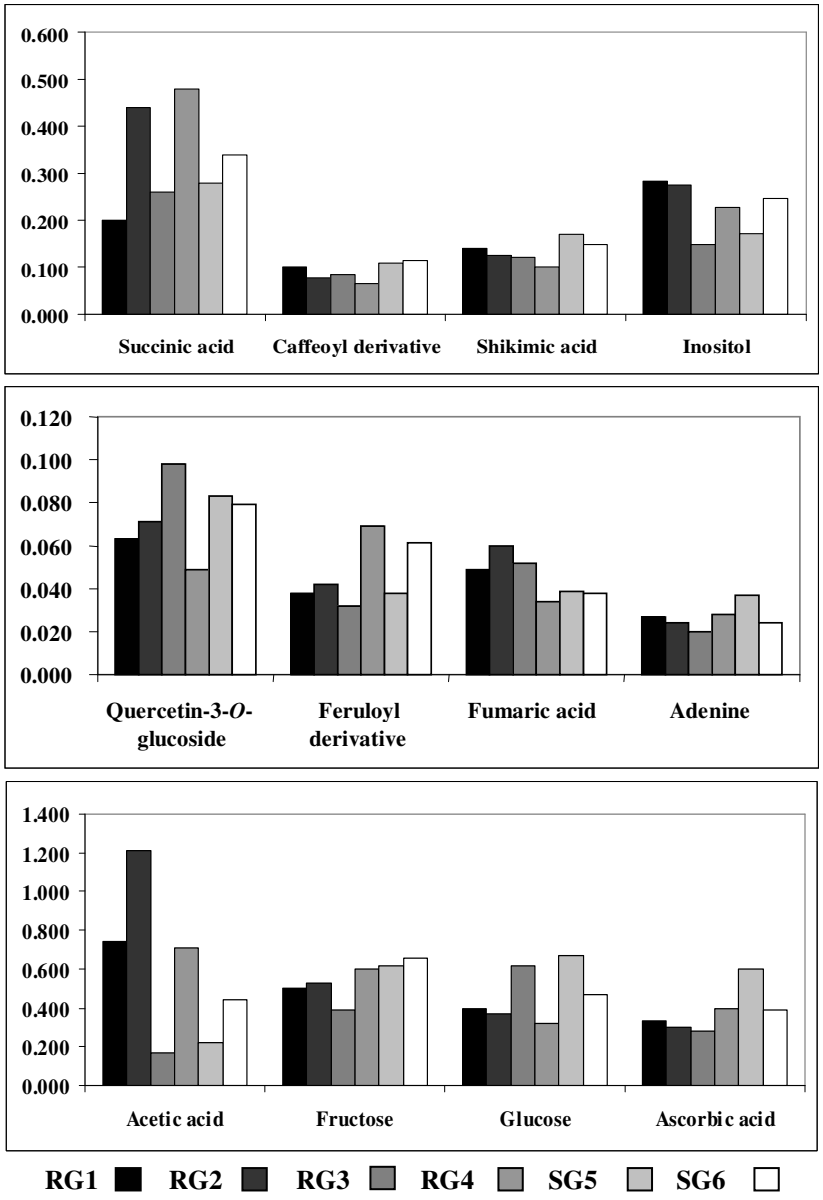

Figure 4. Relative quantification of compounds based on the mean peak area of the signals associated to that compound. Chemical shifts (δ) used for the relative quantification for the compounds are acetic acid at δ 1.94; adenine at δ 8.10; fructose at δ 4.12; glucose at δ 5.17; inositol at δ 3.20; ascorbic acid at δ 4.52; fumaric acid at δ 6.60; succinic acid at δ 2.53; caffeic acid at δ 7.12; shikimic acid at δ 2.20; quercetin-3-*O*-glucoside at δ 6.32; *trans*-feruloyl derivative at δ 7.26. The graph shows the *p* value after *t*-test between resistant and susceptible groups.

Discussion

In plants primary metabolites such as amino acids, nucleotides and carbohydrates, are not only involved in their survival due to the crucial role they play in growth, reproduction and energy generation, but also in the resistance against pathogens (Lokvam et al. 2006), insects (Berenbaum 1995), and herbivores (Rostas et al. 2002). Several primary metabolites were identified in this study including methionine, proline, valine, threonine, alanine, glutamine, glutamic acid, adenine, glucose, sucrose, and fructose. As discussed earlier, many of these primary metabolites are responsible for the metabolic discrimination of grapevine cultivars.

Plants have a very unique property known as genomic plasticity which can be defined as their ability to diversify the defense response against diverse abiotic and biotic stresses. Since plants are sessile, the major strategy employed to combat these stresses, including water deprivation, salinity, nutritional deficiency, intense insolation, adverse climatic conditions, pollutants, pathogens, insects, and phytophags, is the production of phytochemicals generally known as phytoalexins (Harborne 1999). These important phytochemicals are mainly secondary metabolites since they are not directly involved in basic processes of plants such as growth, development, and reproduction but rather function in plant ecological networks (Harborne 2001).

The majority of the phenolic compounds in plants are produced by the phenylpropanoid pathway and these compounds intensely affect plant growth and development along with playing various important roles in many aspects of plant physiology. Examples of these biological functions are the formation of the cell wall polymer lignin from the phenylpropanoid precursors (Li et al. 2008), anthocyanins as floral pigments that attract pollinators (Tanaka et al. 2008), resistance against microbes (Shadle et al. 2003), and flavor and scent compounds derived from phenylpropanoids (Schwab et al. 2008). Moreover valuable bioplastic materials can be made from phenylpropanoids (van Beilen and Poirier 2008). In the present study, many phenylpropanoids such as *cis*- and *trans*-feruloyl derivatives, together with quercetin-3-*O*-glucoside, were identified. As mentioned earlier, these compounds influenced the clustering of different grapevine varieties in the multivariate data analyses.

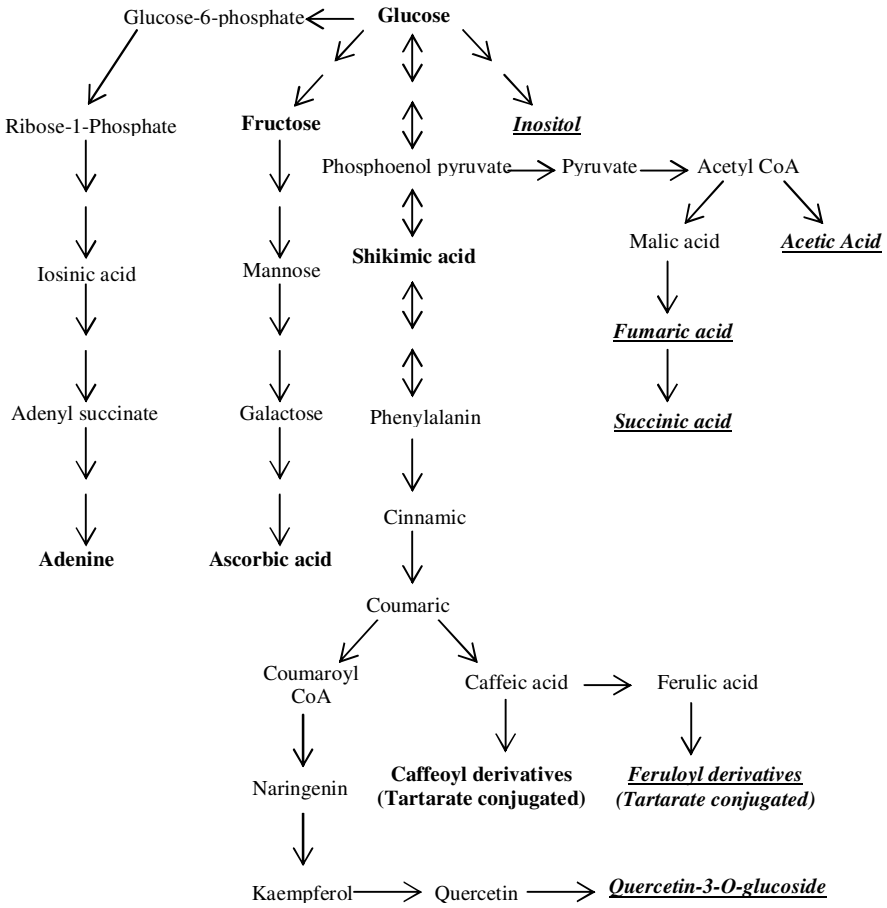

Figure 5. Biosynthetic pathways of the compounds identified in this experiment. Compounds in *italics* with underline show higher level in resistant varieties but compounds in Bold are higher in susceptible varieties.

Conclusion

^{1}H NMR spectroscopy has proved to be an important tool for unbiased metabolite fingerprinting of grapevines. Among several multivariate data analyses, principal component analysis (PCA), hierarchical cluster analysis (HCA), and projections to latent structures-discriminant analysis (PLS-DA) exposed genuine differences between cultivars while the loading plots afforded clues on the nature of this differentiation. Comparison of the spectra of analyzed varieties to a library of NMR spectra of standards run under identical conditions allowed the identification of compounds

responsible for the differences which were observed in both the carbohydrate and the aliphatic regions, including sugars, organic acids and amino acids. In view of these results, it can be easily concluded that resistant varieties exhibit a higher production of many compounds (Figure 5), among them quercetin-3-*O*-glucoside and a *trans*-feruloyl derivative which may contribute to the resistance of these varieties towards Downy mildew. This work shows how ^{1}H NMR analysis can be used for the rapid determination and differential characterization of plant samples based on their metabolic composition. The technique applied here is highly reproducible and covers a wide range of the metabolome. An approach based on hyphenation of the less sensitive NMR with more sensitive methods, such as GC- or LC-MS, for identification of differences in minor compounds seems to be a rational way forward to initiate the screening of plant samples. In this respect we predict many uses of the mentioned NMR technique, from the large-scale analysis of natural variations to the identification of mutants and transgenic plants.