J Am Oil Chem Soc (2011) 88:1463–1475
DOI 10.1007/s11746-011-1812-1

ORIGINAL PAPER

Multivariate Analysis of ^{1}H-NMR Spectra of Genetically Characterized Extra Virgin Olive Oils and Growth Soil Correlations

Paride Papadia · Laura Del Coco · Innocenzo Muzzalupo · Marco Rizzi · Enzo Perri · Gianluigi Cesari · Vito Simeone · Donato Mondelli · Francesco P. Schena · Francesco P. Fanizzi

Received: 15 June 2010 / Revised: 10 February 2011 / Accepted: 15 March 2011 / Published online: 7 April 2011

Abstract Mono-varietal extra virgin olive oils were micro-extracted from drupes that were selectively collected from 28 trees distributed in five different Southern Italian Apulian areas. Nuclear Magnetic Resonance (NMR) profiles of these oil samples were correlated to the genetic (young green material) and soil (samples collected within the foliage projection) data of the tree of origin. Genetic analysis, performed on the samples using SSRs (Simple Sequence Repeats) by 9 microsatellite loci, confirmed the specific cultivar assignment (among Cima di Mola, Coratina, Ogliarola, and Oliva Rossa cultivars). Chemometric methods applied to ^{1}H-NMR spectroscopic data were used for cultivar and geographical origin discrimination of the studied extra virgin olive oils. Linear Discriminant Analysis (LDA) afforded a high reliability degree for discriminating cultivars (almost 90% of prediction ability), and a good assigning ability for the geographical origin (Ogliarola and Coratina samples used as subsets). Soil analyses were performed for each tree. Regression analysis was applied to soil composition in order to correlate available nutrients and total metals with the content of fatty acids and minor components present in monovarietal extra virgin olive oils. In the case of oleic and linoleic fatty acids, and for some terpenes, B, Cr, Mn, Zn were found to give significant correlations. Zn and Mn were the most significant trace elements for all the correlations found ($p < 0.01$). The results obtained (genetic, spectroscopic and soil analyses) are discussed as a multidisciplinary approach for setting up a strategy for a cultivar and/or geographic origin certification committed database construction.

Keywords NMR spectroscopy · Extra virgin olive oil · Statistical analysis · Food origin characterization · Soil analysis

Abbreviations

EEC	European Economic Community
CDM	Cima di Mola cultivar
CTAB	Cetyltrimethylammonium bromide
COR	Coratina cultivar
DTPA	Diethylenetriaminepentaacetic acid
EDTA	Ethylenediamine tetraacetic acid
HCA	Hierarchical clustering analysis

Electronic supplementary material The online version of this article (doi:10.1007/s11746-011-1812-1) contains supplementary material, which is available to authorized users.

P. Papadia · L. Del Coco · F. P. Fanizzi (✉)
Dipartimento di Scienze e Tecnologie Biologiche ed Ambientali, Università del Salento, Prov.le Lecce-Monteroni, 73100 Lecce, Italy
e-mail: fp.fanizzi@unisalento.it

I. Muzzalupo · E. Perri
CRA, Centro di Ricerca per l'Olivicoltura e l'Industria Olearia, C/da Li Rocchi Vermicelli, 87036 Rende (CS), Italy

I. Muzzalupo
CRA, Centro di Ricerca per le produzioni Foraggere e Lattiero-Casearie, Via A. Lombardo 11, 26900 Lodi, Italy

M. Rizzi · F. P. Schena · F. P. Fanizzi
C.A.R.S.O. Consortium, Strada Provinciale Casamassima Km 3, 70010 Valenzano (Bari), Italy

G. Cesari · V. Simeone
C.I.H.E.A.M Istituto Agronomico Mediterraneo di Bari, Via Ceglie 9, 70010 Valenzano (Bari), Italy

D. Mondelli
Dipartimento di Biologia e Chimica Agro-Forestale ed Ambientale, Università di Bari Via E, Orabona 4, 70126 Bari, Italy

LDA	Linear discriminant analysis
NMR	Nuclear magnetic resonance
OGL	Ogliarola cultivar
OLR	Oliva Rossa cultivar
PCA	Principal component analysis
PDO	Protected designation of origin
PVP-40	Polyvinyl-pyrrolidone
SSR	Simple sequence repeat
TEA	Triethanolamine
UPGMA	Unweighted pair group method with arithmetic mean

Introduction

Reference of foodstuffs to the district in which they are produced and to the raw materials used, is an evergreen interest for consumers and likewise for the producers. This connection is needed in order to make such products certified and authenticated, vouching for their differences in taste and nutritional properties from all other regional and/or country products (typicality). The Protected Designation of Origin (PDO) trademark has been assigned to numerous local products based strictly on their area of origin. In order to obtain this designation, the raw materials must have been produced and processed in the specific region from which the product gets its name [1]. By European Union (EU) definition, "PDO products are most closely linked to the concept of *terroir*—a sense of place—discernible in the flavor of the food. PDO products must be produced, processed and prepared in a specific region using traditional production methods. The raw materials must also be from the defined area whose name the product bears. The quality or characteristics of the product must be due essentially or exclusively to its place or origin, i.e. climate, the nature of the soil and local know-how" [2]. The EEC regulation No. 1019/02, modified on March, 6th 2009 [3, 4], concerning the obligation to indicate on the label the origin of the virgin and extra virgin olive oil, is an important step in protecting quality and transparency, because it gives consumers the opportunity to distinguish the products of different countries both among and outside EU countries. The measure has been applied since July, 1st 2009. The typical characteristics of a PDO extra virgin olive oil are the cultivar used, the soil and climate factors affecting the olive tree groves and the production technologies (including agronomical systems, culture practices, soil and plant treatments, irrigation, fertilization, different oil extraction methods). Apulia, Sicily and Calabry are to date the Southern Italian regions responsible for 90% of the olive oil production in Italy. About 5×10^8 kg of olive oil was produced during the 2009/10 season with a slight decrease (−15%) in respect to the previous year (about 6×10^8 kg). This production was compensated for by improved quality, according to the producer organizations [5]. In Apulia, the extra virgin olive oil production is very diverse, due to several environmental factors (differences in microclimatic, pedoclimatic and ecological conditions) present in the same region and various cultivars simultaneously well-cultivated in the same areas surrounding the Apulia provinces. In recent years several attempts were made to define olive oil origin and variety by analysis of chemical parameters [6]. Many analytical techniques were used in order to obtain a suitable method for classification [7], and among them, ^{1}H-NMR spectroscopy [8]. As the concentrations of compounds are of primary interest for food chemistry in the classical assessment procedure, a chemical characterization of olive oil samples was performed, using signals selected in ^{1}H-NMR spectra, in order to calculate the fatty acid composition [9]. Thus, NMR spectroscopy provides a clear advantage in methodology since it is possible to quantify triglyceride percentage and other classical oil parameters (iodine number, peroxide number, etc.). Indeed, the unsaturation degree was evaluated for each cultivar examined, according to Johnson and Schoolery [10]. These results were comparable with conventional analysis, performed by iodometric titration [11]. Although NMR spectroscopy was quite successfully used in order to characterize extra virgin olive oils according to geographical origin and cultivar composition, to date [8, 12] no systematic attempt to obtain a complete monovarietal NMR reference database in order to derive the composition of an unknown olive oil sample has been performed. Herein we report the first attempt to classify the different cultivars from the Apulia region according to genetic identification of the tree, physico-chemical analysis of the plant soil, ^{1}H-NMR spectroscopic data of extra virgin olive oils, and statistical analysis [13]. Systematic investigation of such correlations were applied both in the production and in the optimization of multivarietal extra virgin olive oils, in order to improve the traceability of the raw material used to produce this valuable food.

Materials and Methods

Olive and Plant Material Collection

Olive trees were chosen on the basis of the phenotype, selecting plants which, according to farmers' declarations, had not been grafted. Moreover, young leaves were harvested. All plants were labeled for subsequent identification. Samples were collected during the 2007/08 campaign, from November to December 2007. Fresh plant leaves and

drupes were collected by hand from the top of the olive tree foliage. Coratina (COR), Cima di Mola (CDM), Ogliarola (OGL), Oliva Rossa (OLR) were the olive varieties selected for this study. The production area examined was the Bari district, which includes the following subareas: Northern Coastal area (towns: Bisceglie, Molfetta, and Giovinazzo), Basin area (towns: Bitonto, Palo Del Colle, Bitetto, Binetto, Grumo Appula, and Toritto), North Premurgiana area (towns: Corato, Ruvo di Puglia, Terlizzi, and Andria), Murgiana area (towns: Minervino Murge and Canosa di Puglia), Coastal and South Premurgiana area (towns: Turi, Conversano, Polignano, Monopoli, Castellana Grotte, Putignano, Alberobello, and Locorotondo).

Genetic Analysis

In order to confirm the phenotype attribution of the samples and to build suitable predictive functions with a cross-verified class attribution, genetic analysis was performed a priori.

DNA Extraction

Young olive leaves were washed with 4% of sodium hypochlorite and 0.2 g of leave tissue was ground in a mortar and pestle in 1 mL of warm (60 °C) extraction buffer [100 mM Tris–HCl (pH 8.0), 1.4 M NaCl, 20 mM EDTA (pH 8.0), 2% PVP-40, 2% CTAB], with 1% of 2-mercaptoethanol added 30–60 min before use. The product was transferred in a 5-mL centrifuge tube and the mortar was rinsed with an additional 1 mL of extraction buffer. The mixture was then incubated at 60 °C for 30 min, with agitation approximately every 10 min. After the mixture was cooled at room temperature, an equal volume of dichloromethane/isoamyl alcohol (24:1) was added, and centrifuged at 1,600×*g* for 10 min. The upper (aqueous) phase was transferred in a clean 5-mL centrifuge tube and centrifuged at 1,600×*g* for 10 min. If the aqueous phase was cloudy, the dichloromethane/isoamyl alcohol extraction was repeated. Once clear, the aqueous phase was transferred to a clean tube and two volumes of cold ethanol (−20 °C) and 1/10 volume of ammonium acetate 10 M was added, and the liquid was mixed gently. Samples were left to stand at −80 °C for 30–60 min, or alternatively at −20 °C overnight, and centrifuged at 5,000×*g* for 30 min to pellet the DNA. The pellet was washed with 2 mL of 76% cold ethanol, dried at room temperature overnight and resuspended in 0.1 mL of Tris buffer (10 mM Tris–HCl, pH 8.0). Samples were stored at 4 °C for short-term storage or −20 °C for longer. The total DNA was digested with 1 μL RNase A (stock 50 μg/μL, Sigma-Aldrich Srl, Milano, Italy) and incubated at 37 °C for 30 min. The DNA was extracted with an equal volume of dichloromethane/isoamyl alcohol (24:1) and precipitated with 2 volumes of cold ethanol; it was then dissolved in sterile water. DNA was quantified by H33258 dye incorporation detected by a Hoefer DyNA Quant 200 fluorometer (Amersham Pharmacia Biotech, Milan, Italy) following the method described elsewhere [14]. All chemicals were analytical grade (Sigma-Aldrich Srl) and used without further purification.

Microsatellite Markers

Simple sequence repeat (SSR) amplification was performed by a Gene Amp PCR System 9600 (PE Applied Biosystems, Foster City, CA, USA) in a 25-μL reaction mix containing 20 ng of genomic DNA, 10 mM Tris–HCl pH 8.0, 1.5 mM $MgCl_2$, 50 mM KCl, 0.2 mM dNTPs, 0.25 μM forward and reverse primer and 0.5 U Taq polymerase (Roche Diagnostics, Mannheim, Germany). The thermal profile adopted for GAPU59, GAPU71A, GAPU71B, GAPU103A, UDO012, UDO28, UDO39 primers was as follows: 1 cycle of initial denaturation of 3 min at 94 °C, followed by 40 cycles consisting of 30 s at 94 °C (denaturation), 40 s at 60 °C (annealing) and 30 s at 72 °C (elongation), and finally 1 cycle of final elongation of 5 min at 72 °C. The thermal profile used for primer DCA9 was as follows: 1 cycle of initial denaturation of 5 min at 95 °C, followed by 35 cycles consisting of 20 s at 95 °C (denaturation), 55 s at 30 °C (annealing) and 30 s at 72 °C (elongation), then a final elongation cycle of 5 min at 72 °C. The thermal profile used for primer DCA18 was the following: 1 cycle of initial denaturation of 5 min at 95 °C, followed by 35 cycles consisting of 20 s at 95 °C (denaturation), 50 s at 30 °C (annealing) and 30 s at 72 °C (elongation), then a final elongation cycle of 5 min at 72 °C. Nine published SSR markers were pre-selected for their high level of polymorphism and easy scorable patterns [14, 15]. The used SSRs are reported in Table 1. PCR products were analyzed using a Bioanalyzer 2100 (Agilent Technologies, Waldbronn, Germany) on a DNA 500 LabChip [14]. DNA extractions and PCR amplifications were carried out in triplicates.

Genetic Data analysis

Data were processed using POPGENE 32 software [16]. The software allowed the calculation of the number of alleles and their frequency. The heterozygosity, both observed (H_o) and expected (H_e), was calculated using the same software. The H_e of each locus was calculated according to the formula $H_e = n\left(1 - \sum p_i^2\right)/(n-1)$, where p_i is the frequency of the ith allele and n is the number of gene copies in the sample for the given locus [17]. Deviations of H_o values from Hardy–Weinberg expectations were analyzed using the program Genepop 3.4

Table 1 List of primer sequences used for SSR marker amplification

Locus	Primer sequences (5′ → 3′)
GAPU59	CCCTGCTTTGGTCTTGCTAA
	CAAAGGTGCACTTTCTCTCG
GAPU 71A	GATCATTTAAAATATTAGAGAGAGAGA
	TCCATCCATGCTGAACTT
GAPU 71B	GATCAAAGGAAGAAGGGGATAAA
	ACAACAAATCCGTACGCTTG
GAPU 103A	TGAATTTAACTTTAAACCCACACA
	GCATCGCTCGATTTTTATCC
UDO12	TCACCATTCTTAACTTCACACCA
	TCAAGCAATTCCACGCTATG
UDO28	CTGCAGCTTCTGCCCATAC
	GCAGATCATCATTTGGCACT
UDO39	AATTACCATGGGCAGAGGAG
	CCCCAAAAGCTCCATTATTGT
DCA9	AATCAAAGTCTTCCTTCTCATTTCG
	GATCCTTCCAAAAGTATAACCTCTC
DCA18	AAGAAAGAAAAAGGCAGAATTAAGC
	GTTTTCGTCTCTCTACATAAGTGAC

[18]. To determine significance levels for this test, sequential Bonferroni adjustments were used. The probability of null alleles was estimated according to the formula: $r = (H_e - H_o)/(1 + H_e)$ [19]. The SSR loci discrimination power was calculated according to Brenner and Morris [20]. The alleles detected for each microsatellite were recorded into a data matrix of presence (1) and absence (0) of bands (each allele representing a band). The estimations of genetic similarity based on calculation of Dice [21] similarity coefficients among 28 olive trees were analyzed using NTSysPc program version 2.02 [22]. Finally, a tree was inferred using the unweighted pair group method using an arithmetic average (UPGMA) clustering algorithm. The cophenetic correlation coefficient was calculated, and Mantel's test was performed to control the goodness of fit of a cluster analysis for the matrix on which it was based [23].

Soil Sampling and Analysis

For every plant, soil sample was obtained according to the following procedure: three soil samples were collected assuming the external limit of the foliage projection of each tree (as the right point of sampling). Soil sampling points were approximately located at the tips of an equilateral triangle. For every sampling point the superficial layer was removed about for 5 cm and the sampling was performed from 5 to 50–60 cm in depth, depending on the presence of rocks in the soil. The sampling was performed using a hoe, and the stones were recovered as well, in order to have also a frame determination. Every sample (about 1.5 kg) was labelled with the tree code, and stored for later analysis in the Soil and Water Testing Laboratory of C.I.H.E.A.M. Soil organic carbon content was analyzed according to the Walkley and Black method described by Greweling and Peech in [24]. The method consisted of the measurement of organic carbon through the oxidation of organic carbon compounds by $Cr_2O_7^{2-}$ and subsequent determination of unreduced $Cr_2O_7^{2-}$ by oxidation–reduction titration with Fe^{2+}. Briefly, 1 g of dry fine grained soil (<0.2 mm) was weighed in a conical flask (500 mL). Ten mL of potassium dichromate $K_2Cr_2O_7$ solution (1 N) and 20 mL of concentrated sulfuric acid (H_2SO_4) was added drop by drop. The flask was swirled to mix the suspension and then allowed to stand for 30 min. After that, 200 mL of distillated water was added to cool the solution and to stop the reaction. The titration was done using a solution of iron sulfate heptahydrate $FeSO_4{\cdot}7H_2O$ (0.5 M) after adding *o*-phenanthroline as indicator. The organic matter (OM) content was calculated by multiplying the determined organic carbon by 1.724. Total nitrogen was analyzed using the routine method "Kjeldahl method" [24]. Briefly, 2.5 g of soil was weighed in a glass tube, a teaspoonful of catalyst mixture (K_2SO_4:$CuSO_4$ = 9:1, w/w) was added, then 20 mL of concentrated sulphuric acid (H_2SO_4) were added. Ten mL of hydrogen peroxide (H_2O_2) were added drop by drop and swirled carefully. Then, the sample was digested for 2–3 h at temperature ≈385 °C until the color of the digest became light. The glass tube, after cooling, was inserted into a distiller and 80 mL of sodium hydroxide (NaOH, 40%) were added; thus, a steam distillation was done for 10 min. The distillate was collected in a 500 mL conical flask containing 25 mL of boric acid (H_3BO_3, 2%) with an indicator (mixture of bromocresol green and methyl red in ethanol). At the end the titration was done with 0.1 N H_2SO_4 (normex) until the green color of liquid turned to purple. Total carbonates were measured using the gas volumetric method and the Dietrich-Frühling calcimeter [24]. Briefly, 1 g of dry fine grained soil (<0.2 mm) was weighed in the special flask of the Dietrich-Frühling calcimeter. An excess solution of HCl (1:1, v/v) was inserted into the sample flask, the system then was well closed to the atmosphere and then the soil sample and acid were shaken. The volume of CO_2 produced and the temperature were registered at the time of analysis. Available phosphorus was measured by the Olsen method [24]. Briefly, 5 g of soil and 0.5 of carbon black (to block and remove the dissolved organics) were shaken with 100 mL of $NaHCO_3$ (0.5 N adjusted to pH 8.5) for 30 min. Then, the extract was filtered using a Whatman filter paper No. 42. Phosphorus was measured in a small aliquot of the filtered extract with a Megatech SP 930 spectrophotometer model (absorbance at 665 nm), using the modified ascorbic

acid method. The exchangeable cations were measured using a solution of barium chloride and triethanolamine (TEA) adjusted at pH 8.2 (100 g of $BaCl_2 \cdot 2H_2O$ with 22.5 mL of TEA adjusted at pH 8.2 using HCl 1 M). Briefly, 2.5 g of soil was shaken with 50 mL of the barium chloride solution for 30 min. Then, the sample was filtered using a filter paper Whatman No. 42. The leachate was diluted 1:25, and the exchangeable cations (Ca, K, Mg and Na) concentrations were determined by the Inductively couple plasma optical emission spectrometer (ICP-OES) (Thermo Electron ICAP 6000 Series). For the dosage of available micronutrients a diethylenetriaminepentaacetic acid (DTPA) extraction method was used to measure Zn, Fe, Mn and Cu as described by Lindsay and Norvell in [24]. The extract consisted of 0.005 M DTPA, 0.1 M triethanolamine, and 0.01 M $CaCl_2$, buffered at pH 7.3. Briefly, 50 g of air-dry soil was mixed with 100 mL of extract for 2 h. The sample was filtered, and the micronutrient (Zn, Fe, Mn and Cu) concentration was measured by the ICP-OES (Thermo Electron ICAP 6000 Series). Total cations and heavy metal concentrations were measured by the wet digestion of the air dried sample using the microwave digester (CEM model, MARS Xpress). The sample was mixed with H_2O_2 (1 mL), HCl (1 mL), and HNO_3 (5 mL) for 20 min at 190 °C. The samples were cooled, collected in a flask, and finally measured for their metal content by the ICP-OES (Thermo Electron ICAP 6000 Series).

Monovarietal Extra Virgin Olive Oil Extraction

Monovarietal extra virgin olive oil samples were obtained using the Spremoliva® milling machine [Toscana Enologica Mori, Tavarnelle Val di Persa (FI), Italy]. Each sample was obtained from 25 kg of olives collected from a single tree. Extra virgin olive oil samples were then stored in sealed dark glass bottles labelled with the tree code and kept at room temperature in the dark prior to analysis at the CA.R.S.O. Consortium. The machine was able to process 10 kg of olives per cycle. The duration of a whole working cycle was about 1 h. The cleaning of the machine was necessary after every milling.

NMR Spectra Acquisition

NMR olive oil samples were prepared by mixing 250 μL of the oil sample with 740 μL of deuterated solvent ($CDCl_3$). NMR spectra were obtained at 25 °C by the Bruker Avance DRX 500 spectrometer equipped with an inverse triple resonance, z-gradient probe. For each olive oil sample a 1D ^{1}H-NMR spectrum was acquired. Spectra were obtained by the following conditions: zg pulse program, 32 K time domain, spectral width 14.0019 ppm, p1 8.82 μs (90° is 9.65 μs), pl1 4 db, 64 repetitions, *T* 25 °C.

Characterization of the Extra Virgin Olive Oil by ^{1}H-NMR Analysis

Resonances were assigned on the basis of the literature data [25, 26]. Integrals of selected signals in ^{1}H-NMR spectra were used in order to calculate fatty acid composition for each monovarietal olive oil sample. Intensities were normalized assuming the area of the *sn*-1,3 signal of triacylglycerol as internal standard [9]. Also, ten signals due to minor components of extra virgin olive oils were obtained by ^{1}H-NMR spectroscopy. The 10 resonances chosen for statistical analysis, and selected according to published literature, are due to the following olive oil minor components: hexanal, *trans*-2-hexenal, two other unsaturated aldehydes, another aldehyde, three terpenes, squalene and *β*-sitosterol [6, 26]. A sample extra virgin olive oil ^{1}H-NMR spectrum, with expansions relative to the ten selected resonances is shown in Fig. 1. The intensity of signal was referenced to the resonance at 1.56 ppm, due to *β*-carboxyl protons of all acyl chains, normalized to 1,000 (data in Table S1).

Statistical Analysis

Multivariate statistical analysis and graphics were obtained using the Open Source statistical package R, version 2.10.1 [27]. One-way analysis of variance (ANOVA) and Tukey's multiple comparison test were performed on the fatty acid composition data calculated on the basis of ^{1}H NMR (see Table S2). The ANOVA test is used to assess whether the means are significantly different, while the Tukey test compares a posteriori differences among means with appropriate adjustment for the multiple testing [28]. The ANOVA and Tukey tests were performed in order to evaluate whether the fatty acid composition was correlated with the plant cultivar. For multivariate statistical analysis of minor components data, three different procedures were used: Hierarchical cluster analysis, principal component analysis, and finally linear discriminant analysis. Hierarchical cluster analysis (HCA), a method of unsupervised learning of a dataset partitioning into classes or categories, consisting of elements of comparable similarity, and principal component analysis (PCA), making an a priori unspecified number of clusters, were used. Then, discrimination between samples was done using the linear discriminant analysis (LDA). We aimed at creating a classification of samples with the same characteristics of cultivar and/or geographical origin [13]. Sample code, source, cultivar, and geographic area are reported in Table 2. HCA and PCA were used for an evaluation of the dataset applying two widely used unsupervised classification methods, moreover allowing comparison with the results obtained with a supervised method (LDA) using

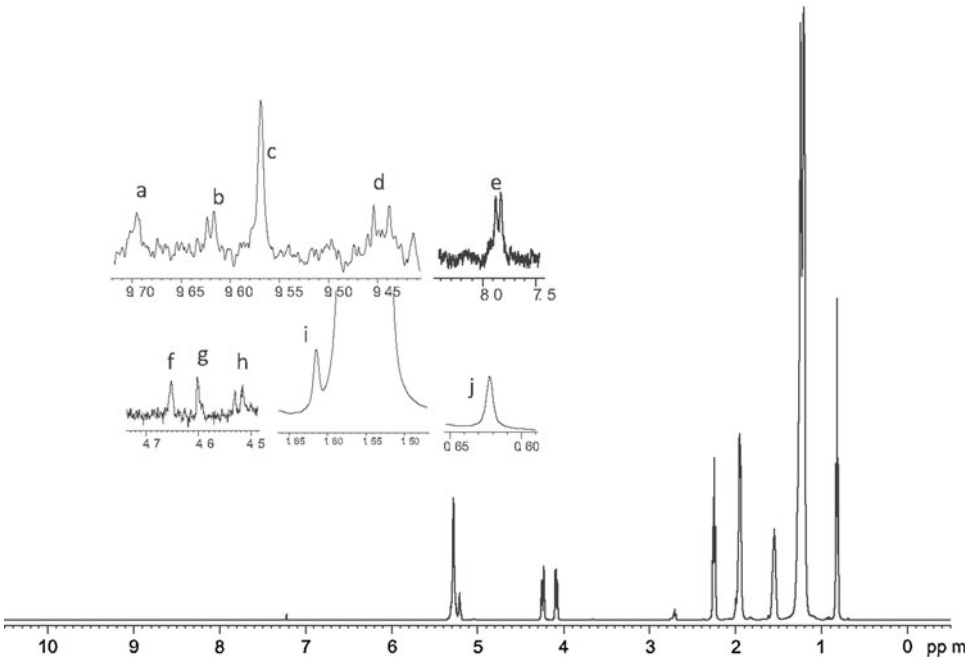

Fig. 1 ^{1}H-NMR spectrum of the extra virgin olive oil OIL1 (cultivar Ogliarola, solvent $CDCl_3$). Signals of minor components used for multivariate analysis are shown in expanded scale and labeled (signals are not at the same scale). **a** hexanal (9.70 ppm); **b**, **c**, **d** unsaturated aldehydes (9.62, 9.57, 9.45 ppm); **e** aldehyde (7.85 ppm); **f**, **g**, **h** terpenes (4.65, 4.60, 4.57 ppm); **i** squalene (1.62 ppm); **j** β-sitosterol (0.622 ppm)

genetic analysis as a starting point. In order to evaluate the performance of the LDA, two different criteria were used: the classification ability and the prediction ability. The first is the percentage of the samples of the training set correctly classified by the discriminant functions. The second is the percentage of samples of the test set correctly classified by using the decision functions developed during the training. The training set used for the evaluation of the classification ability was the whole dataset, while the prediction ability was evaluated developing the decision model on a data subset obtained by removing one sample at a time (leave one out procedure, LOO) [29]. Furthermore, multivariate regression analysis was performed on soil sample data, in order to obtain indication about correlations between ^{1}H-NMR spectroscopic data (triglycerides and minor components) and micro- and macro-elements present in the soil. Stepwise regression analysis was performed [30]. The method can be used when the relationship between one dependent variable and more than one independent variables has to be assessed. Stepwise regression analysis automatically eliminates the independent variables having no significant impact on the dependent variable, thereby increasing the reliability of calculation. Briefly, at each stage in the process, after a new variable is added, a test is made to check if some variables can be deleted without appreciably increasing the residual sum of squares. The procedure terminates when the measure is (locally) maximized or when the available improvement falls below a critical value. The main useful variables determined from the samples characterization were: fatty acids (oleic, linoleic, and linolenic acids) and minor components, previously described [26].

Results and Discussion

Genetic Analysis

A total of 35 alleles for 9 loci were identified (Table 3). All microsatellites were polymorphic, except for GAPU71B, which was monomorphic in all the examined plant leaf samples. A mean of 3.9 alleles per locus was amplified, ranging from 2.0 at GAPU71B to 5.0 at GAPU59, GAPU71A and UDO-39 respectively. This is comparable to the number of alleles among olive cultivars reported by Cipriani et al. [31]. The observed heterozygosity (H_o) for the 28 olive trees ranged from 0.500 at GAPU59 to 1.000 at GAPU71B, GAPU103A and UDO28 with a mean value of 0.794. The expected heterozygosity (H_e) ranged from 0.488 at DCA18 to 0.757 at UDO28, with a mean value of 0.635. The probability of occurrence of null allele values ranged from −0.07 at UDO39 to 0.18 at GAPU103A. At six of the eight polymorphic loci, the observed heterozygosity was higher than the expected values under "Hardy–Weinberg" equilibrium. To discriminate genotypes by the molecular markers, the discrimination power (DP) was computed for each SSR locus (except for GAPU71B locus). The discrimination power varied from 0.497 at GAPU103A to 0.645 at GAPU59, GAPU71A, UDO12, UDO39 and DCA18. The combined DP of all loci was 0.9994 which means that the probability of finding two cultivars with the same genotype combination for the eight SSR markers was 1 over thousands, indicating the elevated discrimination of the marker system used. The shortest allele among the all loci was allele 124 bp at GAPU71B, while the longest was 224 bp at GAPU71A (Table 4). The lowest allelic

Table 2 List of samples

Sample	Cultivar	Farm (name)	Area
OIL30	CDM	Barletta C.	Premurgiana South
OIL31	CDM	Caputo A.	Premurgiana South
OIL32	CDM	Marchitelli P.	Premurgiana South
OIL11	COR	DeRobertis G&G	Hollow Bari
OIL12	COR	Sblano F.	Hollow Bari
OIL14	COR	Siragusa S.	Hollow Bari
OIL15	COR	Speranza F.	Hollow Bari
OIL22	COR	Sorice M.R.	Premurgiana North
OIL23	COR	Altamura M.	Premurgiana North
OIL24	COR	Volpe G.	Premurgiana North
OIL25	COR	Terlizzi L.	Murgiana
OIL27	COR	Terlizzi A.	Murgiana
OIL28	COR	Carlone M.	Murgiana
OIL29	COR	Sassi C.	Murgiana
OIL1	OGL	Caputi G.	North Bari
OIL2	OGL	Balestra	North Bari
OIL3	OGL	Laforgia A. c.da Difesa	North Bari
OIL4	OGL	Laforgia A. c.da Favale	North Bari
OIL6	OGL	Sicolo G.	Hollow Bari
OIL7	OGL	Fasano C.	Hollow Bari
OIL8	OGL	Minenna V.	Hollow Bari
OIL9	OGL	Minenna G.	Hollow Bari
OIL10	OGL	Mitolo	Hollow Bari
OIL16	OGL	Giannelli V.	Premurgiana North
OIL18	OGL	Antonelli G.	Premurgiana North
OIL20	OGL	Cipriani G.	Premurgiana North
OIL36	OLR	Palmisano G.	Premurgiana South
OIL39	OLR	Rotolo G.	Premurgiana South

OIL was the laboratory code used for samples. Cultivar codes: Ogliarola (OGL), Coratina (COR), Cima di Mola (CDM), Oliva Rossa (OLR)

Table 3 Simple sequence repeat amplification products of 28 olive trees

Locus	Na	Ne	H_e	H_o	r	HW[a]	PD
GAPU59	5	2.333	0.582	0.500	−0.05	NS	0.645
GAPU71A	5	3.424	0.721	0.929	0.11	NS	0.645
GAPU71B	2	2.000	0.509	1.000	–	–	–
GAPU103A	3	2.662	0.636	1.000	0.18	NS	0.497
UDO12	4	3.853	0.754	0.821	0.04	NS	0.645
UDO28	4	3.901	0.757	1.000	0.12	NS	0.554
UDO39	5	2.256	0.567	0.464	−0.07	NS	0.645
DCA9	4	3.220	0.702	0.893	0.10	NS	0.589
DCA18	3	1.919	0.488	0.536	0.03	NS	0.645
Mean	3.9	2.841	0.635	0.794	0.08	–	0.541

For each locus, the size range in base pairs, the number of alleles (Na), the effective number of alleles (Ne), the expected heterozygosity (H_e), the observed heterozygosity (H_o), the probability of null alleles (r), the probability of the exact ''Hardy–Weinberg" test (HW) and the discrimination power of alleles (PD)

[a] After Bonferroni correction; *NS* nonsignificant, *S* significance ($p < 0.001$)

Table 4 Allele size (bp) and frequencies (in italics) for each simple sequence repeat locus in 28 olive trees

GAPU59	208	212	214	218	222
	0.214	*0.607*	*0.036*	*0.107*	*0.248*
GAPU71A	210	212	214	218	224
	0.268	*0.054*	*0.214*	*0.054*	*0.411*
GAPU71B	124	144			
	0.500	*0.500*			
GAPU103	136	159	170		
	0.232	*0.268*	*0.500*		
UDO012	166	177	182	193	
	0.268	*0.321*	*0.214*	*0.196*	
UDO028	143	182	205	210	
	0.232	*0.304*	*0.268*	*0.196*	
UDO039	164	185	200	205	220
	0.036	*0.625*	*0.107*	*0.357*	*0.196*
DCA09	182	198	206	210	
	0.107	*0.446*	*0.232*	*0.234*	
DCA18	174	181	185		
	0.054	*0.661*	*0.286*		

frequency (0.036) was observed for the allele 214 bp of GAPU59 and for the allele 164 of UDO39 observed in only two plants of 'Oliva Rossa' (Premurgiana South). The allele 181 bp of DCA18 showed the relatively highest frequency (0.661). For the olive trees (belonging to the four cultivars 'Cima di Mola', 'Coratina', 'Ogliarola' and 'Oliva Rossa'), four SSR profiles were obtained. These different profiles were genetically indistinguishable from SSR profile of the four olive plants growing on olive germplasm collection of the CRA-OLI (Cosenza, Italy). All cultivars were readily separated from each other.

Genetic relationships between the cultivars: Olive genotypes were grouped by cluster analysis as shown in the dendrogram (Fig. 2) based on Dice similarity coefficients. Four clusters distinguished individuals at the variety level. Accessions belonging to the same variety clustered together. The first cluster included 12 'Ogliarola' accessions, the second contained 3 'Cima di Mola' varieties, the third contained the 11 'Coratina' accessions and the fourth included 2 'Oliva Rossa' samples.

Chemical Characterization by NMR Spectroscopy

The fatty acid composition and the iodine number of the triglycerides of the extra virgin olive oils were evaluated by ^{1}H-NMR spectral analysis [9, 25]. The data obtained

Fig. 2 Dendrogram of 28 olive tree genotypes from the Apulia region, generated by UPGMA cluster analysis based Dice coefficient [21]

(Table S2) indicated that all the oils had the expected fatty acid composition of the extra virgin olive oil [32], with oleic acid as the most abundant for all the cultivars (Table 5). One-way ANOVA was performed on the four composition variables as the continuous variable and the cultivar as the categorical variable. The analyses gave a p value <0.001 for all the variables, except for the linolenic acid percentage, which gave a p value <0.05, showing significant differences in the olive oils composition for the different cultivars. Therefore, in order to discriminate significantly different means, the Tukey range test was performed. The oils obtained from Coratina drupes contained on average a greater content (80.5%) of oleic acid, and a lower content (12.4%) of saturated fatty acids than the Ogliarola olive oil samples (76.1 and 15.1% for oleic and saturated fatty acids, respectively, with both Tukey tests adjusted p values <0.001). Linoleic acid content was on the average slightly higher in Ogliarola than Coratina (about 7.7 and 6.1%, respectively, Tukey test adjusted p values <0.02). Mean percentages of fatty acid composition for Ogliarola samples always showed greater standard deviations than Coratina ones for a comparable number of samples. Although the Cima di Mola and Oliva Rossa do not have comparable number of samples to the Ogliarola and Coratina cultivars, some observation can be made. The mean fatty acid composition of Oliva Rossa seems to be quite similar to Ogliarola, with no significant difference in any of the composition variables. On the other hand, Cima di Mola cultivar showed both the highest linoleic (11.1%, Tukey test adjusted p values always <0.01) and saturated fatty acid (17.3%, Tukey test adjusted p values always <0.05) content among all the four studied cultivars.

Multivariate Analysis of Monovarietal Extra Virgin Olive Oil NMR Data

Hierarchical Cluster Analysis

As preliminary investigations, HCA and PCA were carried out using minor components as variables, since these components have been suggested as being the most suitable for discrimination purposes [6, 26]. HCA sorts the samples on the basis of calculated multidimensional distances. The result of hierarchical clustering is a dendrogram which provides an easy visual estimation of relationships among samples defined in a multidimensional measurement space. The distances between samples and groups of samples are the measure of similarity. The Ward's method clustering technique and squared Euclidean distance calculation applied on our analytical data gave the best clustering results shown in Fig. 3; [33]. The dataset exhibited a good clustering of the monovarietal olive oils grouped on the basis of their nominal cultivar. Apart from OIL1, which came on its own as a separate sample, we obtained two main clusters, one almost completely composed of

Table 5 Mean percentage of fatty acid composition for each cultivar determined by ^{1}H-NMR spectral analysis (SD = standard deviation)

Cultivar	Oleic acid (%)	SD	Saturated fatty acids (%)	SD	Linoleic acid (%)	Sd	Linolenic acid(%)	SD
Cima di Mola	70.5	1.8	17.3	0.7	11.1	1.0	1.1	0.1
Coratina	80.5	1.3	12.3	0.7	6.1	0.7	1.0	0.5
Ogliarola	76.1	2.9	15.0	1.4	7.7	1.5	1.1	0.1
Oliva Rossa	76.8	0.8	13.7	0.2	8.5	0.6	1.02	0.01

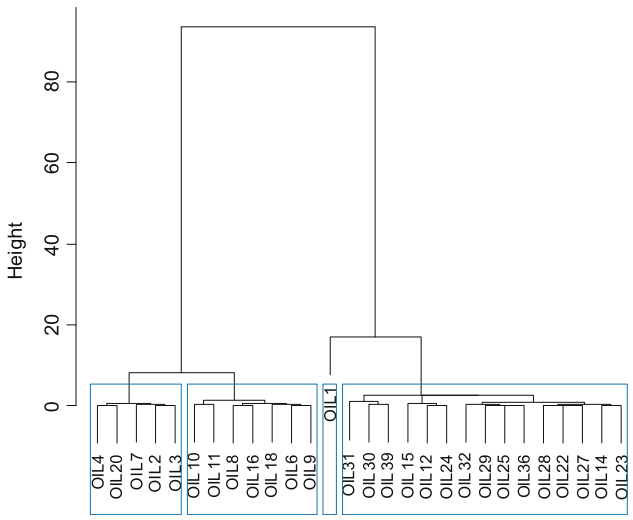

Fig. 3 Cluster dendrogram for the extra virgin olive oil dataset (inputs variables = minor components; method = ward; distance = squared − Euclidean)

Ogliarola samples (11 over 12, with the only exception being Coratina sample OIL11), and the other one containing Coratina, Oliva Rossa and Cima di Mola. This latter can be further partitioned in subgroups, one containing Cima di Mola (2) and Oliva Rossa (1), the second containing only Coratina (3) and the third containing Coratina (8) and Oliva Rossa (1) samples. Interestingly, while the source had practically no influence on the distribution of Coratina and Ogliarola, all samples in the last subgroup come from the same geographical district (Premurgiana South). HCA repeated using the fatty acid percentages calculated by ^{1}H-NMR spectral analysis, and the same clustering criteria, gave satisfactory results, giving good clustering for Coratina samples, but at the same time giving two Ogliarola/Cima di Mola and Ogliarola/Oliva Rossa subgroups (data not shown).

Principal Component Analysis

PCA is a well-known explorative technique to visualize data and to find the true dimension of a data set. The *n* parameters (variables), measured for each sample, describe each sample (object) in an *n* dimensional space. PCA generates a set of new orthogonal variables, linear combinations of the original ones, so that the maximal amount of variance contained in the data set (information) is concentrated in these new variables called principal components (PC) [34]. Generally, variables constituting the PC are reported as a function of each other in a two or three dimensional graph by using two or three PC, respectively. We performed analysis on the dataset in order to discriminate between cultivars. Also, two subsets containing only Coratina samples and Ogliarola samples were separately analyzed on the basis of their geographical origin. The Coratina and Ogliarola samples were all collected in the Bari district, in three different but partially overlapping agricultural areas. The PCA analysis was performed using as parameters the areas of the signals of minor components (see "Characterization of the Extra Virgin Olive Oil by ^{1}H-NMR Analysis" section, integral data are reported in Table S1). The first four components (PCs) explained 96% of the total variance (PC1 65.3%, PC2 18.6%, PC3 6.2%, and PC4 5.5% of variance), while to obtain 99% of the total variance, six components were needed. A scatter plot matrix of the first four PCs is reported in Fig. 4 (parameter loadings in Table S3, PC scores in Table S4). From this plot, a separation of samples according to cultivar was observed. In particular, Oliva Rossa samples clearly separated on PC4, always showing negative scores (lower than −1), whereas all the other samples had positive scores. In the same scatter plot matrix, the separation of Coratina and Ogliarola samples was partially obtained on PC2, where all Coratina samples had scores lower than 0.3 whereas Ogliarola samples had generally higher positive scores but also showed a much greater spread and some outliers. The Cima di Mola samples had PC3 and PC4 always negative and positive respectively but did not show complete separation from Ogliarola and Coratina. PCA did not show a significant distribution in the scatter plot matrix on the basis of the geographical origin of olive oil samples. It should be noted that separation found by simple PCA is rather low, because of the nature of the function used to perform the analysis. In fact, principal components are aligned along the direction of maximum variance, which do not necessarily coincide with maximum intergroup discrimination. The non-supervised HCA and PCA analyses, performed using the minor components of the olive oil showed the distribution of the samples according to the cultivars, in particular for the most represented cultivars in the data set, suggesting the need for a dozen of samples for each cultivar in order to obtain for it a good representation of its typical features. Moreover, we observed that Cima di Mola and Oliva Rossa were characterized as similar by the HCA. On the other hand, we did not observe good discrimination for the sample origin. Qualitatively similar results, but with a lower degree of success, were obtained using the simple fatty acid composition as input data (data not shown).

Linear Discriminant Analysis

In order to fully exploit our approach which starts with genetic analysis we investigated by LDA, a supervised method, using our sample library for predictive purposes. LDA is a classification tool which consists of calculating linear combinations (classification functions) of the original variables able to maximize differences between groups

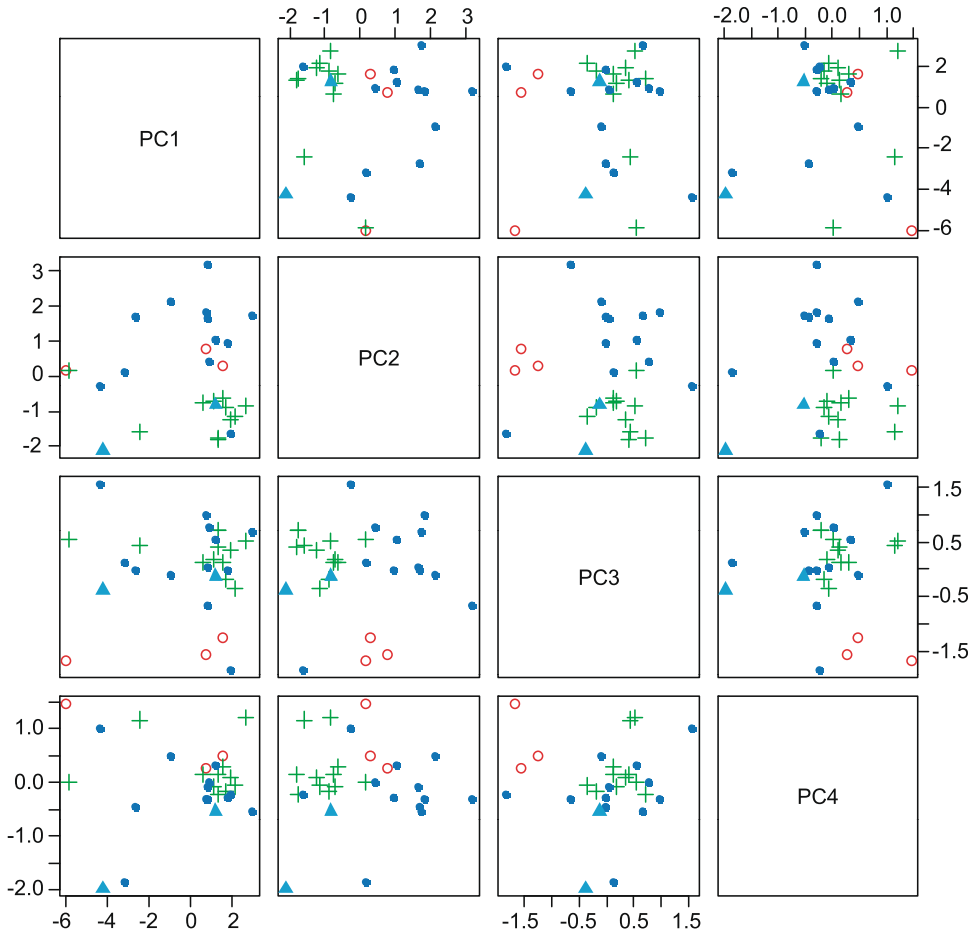

Fig. 4 Scatterplot matrix of the first four PCA components scores for the olive oil data set, grouped according to cultivars (*open circles* CDM; *plus* COR; *closed circles* OGL; *triangles* OLR)

and minimize differences within groups [35]. Two separate LDAs were carried out: one using the fatty acid percentages as input variables (4 variables), and another using minor components content, previously described (10 variables). The cultivar classification ability in the first LDA, was quite good (around 90%), considering that only four variables for each sample were used. The scatter plot matrix of the LD discriminant functions showed tightly (Cima di Mola and Oliva Rossa) broad (Ogliarola) and sufficiently (Coratina) grouped samples (data not shown). The second LDA analysis (see Table S5 for coefficients of linear discriminant), assuming as descriptive variables the minor components of extra virgin olive oil, allowed almost 90% correctness in discrimination of the samples by cultivar. As shown in Fig. 5, samples were correctly assigned, in spite of the low number of Cima di Mola and Oliva Rossa. Moreover, subsets were generated selecting only olive oil samples belonging to the same cultivar, when the number of samples for each area was at least three. Only two cultivars satisfied this prerequisite, Coratina and Ogliarola (11 total samples for Coratina and 12 total samples for Ogliarola). The Ogliarola samples were distributed on three different areas of Apulia (North Bari, Hollow Bari, North Premurgiana), while Coratina samples derived from the areas Hollow Bari, Murgiana, and North Premurgiana. LDA was therefore performed on the two subsets, using the minor components data as input data and geographic origin as discrimination category (see Tables S6 and S7 for coefficients of linear discriminant). The results obtained show that, when analysis is performed on extra virgin olive oils milled from a single cultivar, further discrimination based on the geographical area is possible (Fig. 6). These findings, compared with the predictive ability shown for Ogliarola and Coratina samples, seem to suggest that, for a single year, in order to build up discriminant functions able to distinguish both for cultivars and geographical origin of extra virgin olive oil samples, using only the ^{1}H-NMR spectral data, probably around eight or nine monovarietal samples coming from a single geographical area are needed.

Linear Regression Analysis for Micronutrients and Total Metals

It has already been reported in the literature that the fatty acid content of plants is strongly influenced by soil

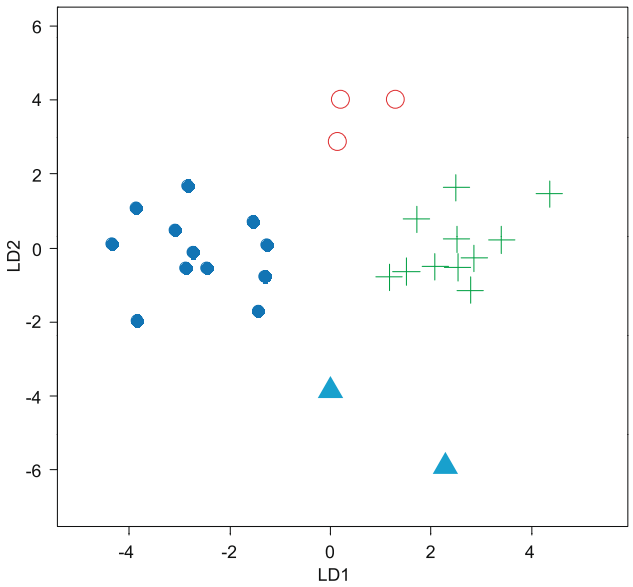

Fig. 5 Scatterplot of the first two LD functions scores grouped according to cultivar (*open circles* CDM; *plus* COR; *closed circles* OGL; *triangles* OLR)

composition and even by the presence of heavy metals. Moreover, metals can act as factors that induce compositional changes of membrane lipids [36]. In fact, cells modify their membrane lipid composition and such changes are thought to help to restore optimal physical properties. Since it is well-known that the oxidative stress arising from such metal exposition could generate reactive oxygen species (ROS) which can interact with polyunsaturated fatty acids (PUFA) and fatty acids, we investigated the possible correlation between extra virgin olive oil FA composition, its minor components (called dependent, output or response variables) in respect to the total metals and the available nutrients (independent or input variables) present in soil. The method used was linear regression analysis, by which the relationship between one or more independent variables and another variable, called dependent variable, is modeled by a least square function, called the linear regression equation. This function is a linear combination of one or more model parameters, called regression coefficients. Since many variables deriving from the soil analyses were available, we looked for possible linear correlations between two or more independent variables as a preliminary step. The correlation matrix revealed that organic matter, total nitrogen, and organic carbon gave redundant information about the soil composition (Pearson's product-moment correlation coefficients were all ≥ 0.99, with a $p < 0.001$). In the same way a relationship was found between available phosphorus and available P_2O_5 (Pearson's product-moment correlation coefficients was greater than 0.99, with $p < 0.001$). Therefore, a subset of the total available micro- and macronutrients data was selected for the subsequent regression analysis. Specifically, nutrients as exchangeable Ca, K, Mg, and Na, total carbonate, total nitrogen, available phosphorous, and metals such as B, Cd, Cu, Cr, Fe, Mn, Ni, Pb, and Zn were used. The regression analysis was performed on a dataset of 19 unique samples, composed by the complete radical soil analysis (Tables S8–S10) and NMR spectroscopic data of the extra virgin olive oil, each sample describing a single tree. The variables were selected according to the method described in the experimental section. A correlation was observed for some of the fatty acids, namely for oleic and linoleic acids (oleic acid showing dependence on B, Cr, Mn, Ni, Zn, adjusted R^2: 0.6991; linoleic acid on B, Cr, Fe, Mn, Zn, adjusted R^2: 0.8041, see supplementary Tables S11–S12), and also among minor components, some of the terpenes (both signals showing correlation on B, Cr, Mn, Zn, Ni, Pb, terpene 1 adjusted R^2: 0.7667; terpene 2, adjusted R^2: 0.6615, see supplementary Tables S13–S14). Interestingly, for both fatty acids and terpenes, significant correlations were observed only with variables relating to total content of the metals present in soil. Regression analysis indicated that, for both oleic and linoleic acid, the most significant variables were Mn and Zn (in all cases $p < 0.001$), with Mn showing a greater significance in the case of linoleic

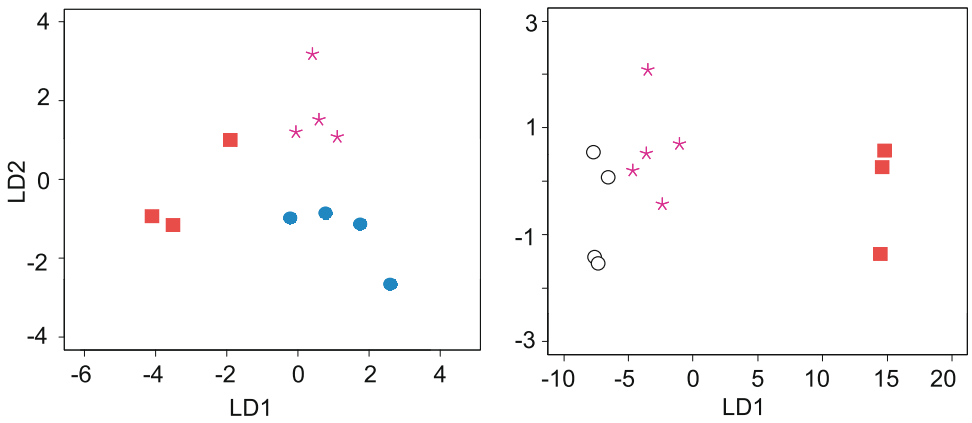

Fig. 6 Scatterplot of the first two LD functions scores for the olive oil subsets, grouped according to geographical origin. (*right*: Coratina samples, *closed circles* Murgiana; *asterisks* Hollow Bari; *closed squares* Premurgiana North; *left*: Ogliarola samples, (*closed squares* Premurgiana North; *open circles* North Bari; *asterisks* Hollow Bari, *closed circles* Murgiana)

acid. In the case of terpenes, both showed the greatest dependence on Mn and Zn (terpene 1, Zn $p < 0.001$, Mn $p < 0.01$; terpene 2, Zn $p < 0.01$, Mn $p < 0.05$) with Zn having in both cases the highest significance. Finally, there were no significant leverages on linolenic acid concentration (C18:3, $\Delta^{9,12,15}$, best correlation obtained with B, Cr, Fe, Mn, Ni, Zn, with adjusted R^2: 0.3348, see supplementary Table S15), perhaps because of its more complex biosynthetic pathway. To sum up, for oleic and linoleic fatty acids and for some of the terpenes contained in the olive oil, the most significant correlations with soil nutrients were found with the total metal concentration. In all cases, B, Cr, Mn, Zn were found to give significant correlation, with Zn and Mn being the most significant trace elements for all the correlations found within olive oil. Moreover, Fe, Ni, and Pb were always found to give correlation in the case of terpenes. On the other hand, in the case of fatty acids, Pb showed no meaningful correlation, with Ni being significant only in the case of oleic acid and Fe in the case of linoleic acid. The lack of correlation of organic nutrients (C, N, P), alkali and alkaline earths metals (Na, K, Ca, Mg) may be due to the fact that the correlation obtained are not connected with the generic welfare of the plant, but with metabolic pathways leading to specific lipidic metabolites contained in the final extracted product, the extra virgin olive oil.

Conclusions

NMR profiles of 28 monovarietal extra virgin olive oils microextracted from olives selectively collected from 28 trees (expected cultivars: Coratina 11, Cima di Mola 3, Ogliarola 12, Oliva Rossa 2) distributed in 5 southern areas of Apulia, were correlated to the genetic (young green material) and soil (samples collected within the foliage projection) analyses for each tree. Genetic analysis was performed using microsatellite markers (SSRs). Linear Discriminant Analysis (LDA), applied to ^{1}H-NMR spectroscopic data, was successfully used for cultivar and geographical origin discrimination of the studied extra virgin olive oils. This method, unlike the unsupervised HCA and PCA, afforded a high reliability degree for discriminating cultivars (ca. 90% of prediction ability). Furthermore, when the analysis was performed on monovarietal samples subsets, a further discrimination (100%), on the basis of the geographical origin, was possible. This result, obtained on Ogliarola and Coratina samples, is particularly interesting for development of PDO assessment, since these two cultivars are the most widespread commercial varieties in the investigated district. Regression analysis was applied to soil composition in order to correlate available nutrients and total metals with the quantity of fatty acids and minor components present in monovarietal extra virgin olive oils. In the case of oleic and linoleic fatty acids, and for some terpenes, B, Cr, Mn, Zn were found to give significant correlation, with Zn and Mn ($p < 0.01$) being the most significant trace elements for all the correlations found. To the best of our knowledge, this is the first time a systematic effort to obtain genetic characterization by SSR amplification, soil analyses, and ^{1}H-NMR spectra, is carried out in order to make a direct connection between the olive tree cultivar (genetic information) and the NMR spectra (chemical information) of the extra virgin olive oil produced. The results herein reported show that a multidisciplinary approach, through the application of multivariate statistical analysis, could be used to set up a method for cultivar and/or geographic origin certification, based on the construction of a suitable database. Further research will be directed to the growth of an organic genetic/NMR/soil database, in order to improve the prediction ability of the LDA, and furthermore to develop a way to correlate ^{1}H-NMR spectra of commercial extra virgin olive oils with their geographical and genetic origin.

Acknowledgments The Authors would like to give special thanks to the olive oil Producers' Association "O.P. Oliveti Terra di Bari", in particular to the President Mr. Gennaro Sicolo, and Coordinator Dr. Raffaele Valentini. P.P. would like to thank Dr. Nicola Zaccarelli for helpful discussions. Contribution from the Italian Ministry of University and Research (MIUR) ex-60% funds, and from the Italian Ministry of Agriculture, Food and Forestry Policies (RIOM-additional grant project) is gratefully acknowledged. C.I.R.C.M.S.B. Interuniversity Consortium (Bari, Italy) provided L.D.C. a studentship grant.

References

1. Martins-Lopes P, Gomes S, Santos E, Guedes-Pinto H (2008) DNA markers for Portuguese olive oil fingerprinting. J Agric Food Chem 56:11786–11791
2. The olive oil source (accessed May 2010) The PDO definition. http://www.oliveoilsource.com/
3. CEE Regulation 1019 (2002) On marketing standards for olive oil. CEE Official Report L155:27–31
4. CEE Regulation 1019 (2002) Amending regulation on marketing standards for olive oil. CEE Official Report L63:6–8
5. ISMEA (accessed May 2010) Ismea: +10% la produzione di olio di oliva in Italia. http://www.ismea.it/flex/cm/pages/ServeBLOB.php/L/IT/IDPagina/3679
6. Sacchi R, Mannina L, Fiordiponti P, Barone P, Paolillo L, Patumi M, Segre A (1998) Characterization of Italian extra virgin olive oils using ^{1}H NMR spectroscopy. J Agric Food Chem 46:3947–3951
7. Mannina L, Dugo G, Salvo F, Cicero L, Ansanelli G, Calcagni C, Segre A (2003) Study of the cultivar-composition relationship in Sicilian olive oils by GC, NMR, and statistical methods. J Agric Food Chem 51:120–127
8. Vlahov G, Del Re P, Simone N (2003) Determination of geographical origin of olive oils using ^{13}C nuclear magnetic resonance spectroscopy. I. Classification of olive oils of the Puglia

region with denomination of protected origin. J Agric Food Chem 51:5612–5615
9. Guillen MD, Ruiz A (2003) Rapid simultaneous determination by proton NMR of unsaturation and composition of acyl groups in vegetable oils. Eur J Lipid Sci Technol 105:688–696
10. Johnson LF, Shoolery JN (1962) Determination of unsaturation and average molecular weight of natural fats by nuclear magnetic resonance (NMR). Anal Chem 34:1136–1139
11. Spyros A, Dais P (2009) ^{31}P NMR spectroscopy in food analysis. Prog Nucl Magn Reson Spectrosc 54:195–207
12. D'Imperio M, Mannina L, Capitani D, Bidet O, Rossi E, Bucarelli FM, Quaglia GB, Segre A (2007) NMR and statistical study of olive oils from Lazio: a geographical, ecological and agronomic characterization. Food Chem 105:1256–1267
13. Giansante L, Di Vincenzo D, Bianchi G (2003) Classification of monovarietal Italian olive oils by unsupervised (PCA) and supervised (LDA) chemometrics. J Sci Food Agric 83:905–911
14. Muzzalupo I, Stefanizzi F, Perri E (2009) Evaluation of olives cultivated in Southern Italy by SSR markers. Hort Sci 44:582–588
15. Baldoni L, Cultrera NG, Mariotti R, Ricciolini C, Arcioni S, Vendramin GG, Buonamici A, Porceddu A, Sarri V, Ojeda MA, Trujillo I, Rallo L, Belaj A, Perri E, Salimonti A, Muzzalupo I, Casagrande A, Lain O, Messina R, Testolin R (2009) A consensus list of microsatellite markers for olive genotyping. Mol Breed 24:213–231
16. Yeh FC, Boyle TJB (1997) Population genetic analysis of codominant and dominant markers and quantitative traits. Belg J Bot 129:157–201
17. Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci USA 70:3321–3323
18. Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86(3):248–249
19. Brookfield JF (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol 5:453–455
20. Brenner C, Morris J (1990) Paternity index calculations in single locus hypervariable DNA probes: validation and other studies. Proc Int Symp Hum Identif Promega Corp pp 21–53
21. Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26:297–302
22. Rohlf M (1998) NTSYS-PC Numerical taxonomy and multivariate analysis system, version 2.02i. Dep Ecol Evol State University NY, Setauket, NY
23. Mantel N (1967) The detection of disease clustering and a generalized regression approach. Cancer Res 27:209–220
24. Swift R (ed) (1996) Methods of soil analysis. Part 3, Chemical Methods. Soil Science Society of America. American Society of Agronomy, Madison
25. Knothe G, Kenar JA (2004) Determination of the fatty acid profile by ^{1}H NMR spectroscopy. Eur J Lipid Sci Technol 106:88–96
26. Mannina L, Segre A (2002) High resolution nuclear magnetic resonance: from chemical structure to food authenticity. Grasas Aceites 53:22–33
27. R Development Core Team (2009) R: a language and environment for statistical computing. Vienna, Austria
28. Yandell BS (1997) Practical data analysis for designed experiments. Chapman & Hall/CRC Press, London
29. Massart DL, Vandeginste BGM, Deming SM, Michotte Y, Kaufman L (1988) Chemometrics: a textbook. Elsevier Science B. V, Amsterdam, pp 385–387
30. Cohen J, Cohen P (1975) Applied multiple regression/correlation analysis for the behavioral sciences. Lawrence Erlbaum Associates New York
31. Cipriani G, Marrazzo MT, Marconi R, Cimato A, Testolin R (2002) Microsatellite markers isolated in olive (*Olea europaea* L.) are suitable for individual fingerprinting and reveal polymorphism within ancient cultivars. Theor Appl Genet 104:223–228
32. Pasqualone A, Montemurro C, Summo C, Sabetta W, Caponio F, Blanco A (2007) Effectiveness of microsatellite DNA markers in checking the identity of protected designation of origin extra virgin olive oil. J Agric Food Chem 55:3857–3862
33. Frank IE, Todeschini R (1994) The data analysis handbook. Elsevier, Amsterdam
34. Brescia MA, Caldarola V, De Giglio A, Benedetti D, Fanizzi FP, Sacco A (2002) Characterization of the geographical origin of Italian red wines based on traditional and nuclear magnetic resonance spectrometric determinations. Anal Chim Acta 458:177–186
35. Rezzi S, Axelson DE, Heberger K, Reniero F, Mariani C, Guillou C (2005) Classification of olive oils using high throughput flow ^{1}H NMR fingerprinting with principal component analysis, linear discriminant analysis and probabilistic neural networks. Anal Chim Acta 552:13–24
36. Bidar G, Verdin A, Garcon G, Pruvot C, Laruelle F, Grandmougin-Ferjani A, Douay F, Shirali P (2008) Changes in fatty acid composition and content of two plants (*Lolium perenne* and *Trifolium repens*) grown during 6 and 18 months in a metal (Pb, Cd, Zn) contaminated field. Water Air Soil Pollut 192:281–291