Chemometric Classification of Olive Cultivars Based on Compositional Data of Oils

Maria Antonietta Brescia[a,*], Giovanna Alviti[b], Vitantonio Liuzzi[b], and Antonio Sacco[a]

[a]Dipartimento di Chimica, Università degli Studi di Bari, 70126 Bari, Italy, and
[b]Dipartimento di Produzione Animale, Università degli Studi di Bari, Facoltà di Agraria, 70126 Bari, Italy

ABSTRACT: European Union regulations provide important guidelines for maintaining the Protected Designation of Origin (PDO) of olive oil and other foods. This includes characterization of foods based on variety (cultivar) and geographical origin, as this may be used as a criterion for determining authenticity and quality. Therefore, analytical method standards need to be established to ensure these criteria. This study describes how cultivar differences can be established between Italian oils, obtained from single varieties, based on acid, sterol, and TAG differences determined by chemometrics. TAG and FA composition provided the best basis for differentiation of olive oils among cultivars. The results were compared with those obtained using ^{13}C NMR analysis, and a similar differentiation between oils of different cultivars was achieved. ^{13}C NMR provides useful information on the acyl composition and on the positional distribution of the glycerol moiety and can be used for classification of cultivars based on oil composition. Furthermore, the advantages of this technique come from the rapidity with which information can be obtained and from the very simple preparation procedure required for analysis.

Paper no. J10268 in *JAOCS 80,* 945–950 (October 2003).

KEY WORDS: Characterization, chemometrics, cultivar, fatty acids, ^{13}C NMR, olive oil, sterols, triacylglycerols.

Virgin olive oil is a food with high nutritional qualities; hence, it is chosen for cooking and dressings by people from Mediterranean countries, which represent the world's largest production area. World olive oil consumption has increased (1) owing to its flavor, health benefits as a monounsaturated oil, and antioxidant content (2). The flavor of extra virgin olive oil has particularly desirable organoleptic and nutritional properties that contribute to its market value. However, olive oil characteristics and compositions are also determined by climate, agronomic factors, extraction methods, and processing techniques that affect the chemistry and flavor of the oil and can vary by growing location. The European Union regulations provide guidelines for maintaining the Protected Designation of Origin (PDO) of olive oil and other foods. This includes characterization of foods based on cultivar and geographical origin, as they are used as an indicator of authenticity and quality. Therefore, there is an economic basis for identifying characteristics that distinguish PDO of olive oils. This has been tackled by many research institutions and has led to the application of different methods to authenticate PDO. FA and sterol compositions have been evaluated as a means of determining the origin of Tuscan oils (3), FA for Greek (4) and Italian (5,6) oils, and TAG for oils from around the world (7).

The objective of this paper is to find a basis for distinguishing among 37 olive oils from the southern Italian region of Apulia according to cultivar. For this purpose, analytical parameters referring to the acidic, sterol, and TAG compositions were evaluated using official methods. These methods often require various stages of sample pretreatment, e.g., separation or concentration. It has recently been shown that ^{13}C NMR spectroscopy enables a direct and fast evaluation of the chemical composition of food. The data obtained were useful in distinguishing the variety and origin of virgin olive oils (8) and in classifying wines (9). ^{13}C NMR was also used for comparative purposes with official methods in this study.

EXPERIMENTAL PROCEDURES

Thirty-seven Apulian virgin olive oils from the 1999–2000 harvest were obtained from the Dipartimento di Produzione Animale, University of Bari, Italy, and from local oil producer associations. Oils were obtained from the Apulia region and included: 7 samples of Coratina, 10 of Leccino, 7 of Peranzana, and 13 of Oliarola.

Analytical determinations. FA as methyl esters obtained after transesterification of the TAG were analyzed by GC as reported previously (6). The individual concentrations (%) of 11 FA (palmitoleic 16:1, heptadecanoic 17:0, heptadecenoic 17:1, linolenic 18:3, stearic 18:0, oleic 18:1, linoleic 18:2, arachidic 20:0, eicosenoic 20:1, behenic 22:0, and lignoceric 24:0) were determined.

The analyses of sterol and TAG fractions were carried out according to official analytical methods (10). Sterol composition was determined after saponification performed on a 5-g sample to which 0.5 mL of α-colestanol as internal standard and 50 mL of a 2 N ethanolic KOH solution were added. The reaction flask, fitted with a reflux condenser, was maintained at gentle boiling until the solution was clear. The mixture was transferred into a separating funnel and extracted with 80 mL of diethyl ether. The ether extract was collected, washed with distilled H_2O, dried with anhydrous Na_2SO_4, filtered, and solvent was removed by rotary evaporation.

*To whom correspondence should be addressed at Dipartimento di Chimica, Campus Universitario, Via Orabona 4, 70126 Bari, Italy.
E-mail: brescia@lgxserve.ciseca.uniba.it

The sterol fraction was separated from the unsaponifiable fraction by TLC with a hexane/diethyl ether 65:35 (vol/vol) mixture as eluting solvent.

The samples were derivatized for GC analysis as trimethylsilyl derivatives. The silylating reagent was a mixture of 100 µL of *N*,*O*-bis-trimethylsilylacetamide and 200 µL of pyridine.

The qualitative and quantitative analyses were performed with an HP 5890A gas chromatograph. The instrument was equipped with a fused-silica capillary column (30 m × 0.25 mm × 0.25 µm) with a poly (5% diphenyl/95% dimethylsiloxane) stationary phase. The carrier gas was helium at a flow rate of 22 cm/s. The temperature of the FID was 320°C. The oven temperature was 270°C, the temperature of the injector (split) was 290°C, and the split ratio was 1:30.

For analysis of the TAG, oil samples were diluted to a 5% concentration in acetone and directly injected into an HPLC chromatograph (Waters 600 E System controller) equipped with a Nova Pak C18 column, 30 cm × 3.9 mm i.d., with a differential refractometer detector (Waters 410) at 40°C and at λ 210 nm. A mixture of acetone/acetonitrile 2:1 was used as eluting solvent at a 0.5 mL/min flow rate. The sterols determined were cholesterol, 24-methylene-cholesterol, campesterol, campestanol, stigmasterol, clerosterol, sitostanol, Δ^5-avenasterol, $\Delta^{5,24}$-stigmastadienol, Δ^7-stigmastenol, Δ^7-avenasterol, β-sitosterol, total sterols, and eritrodiol + uvaol. The TAG that were determined included LLL, OLLn, PLLn, PPoLn, OOLn + PoOL, PLL, POLn, EeOL, OOL, PoOO, POL, PoPO, PPL, PPO, GaOO, SOO, PSO, AOO, and SSO, where L = linoleic, O = oleic, Ln = linolenic, P = palmitic, Po = palmitoleic, Ln = linolenic, Ee = eicosenoic, Ga = gadoleic, S = stearic, and A = arachidic.

NMR determinations. The following procedure for sample preparation for ^{13}C NMR analysis was optimized to obtain high-resolution spectra. Oil samples (5 mL) were dried over anhydrous Na_2SO_4 and placed in an ultrasonic bath for 10 min. After filtration, 750 µL of oil was introduced into a 10-mm NMR tube together with 2.25 mL of $CDCl_3$ (Cambridge Isotope Laboratories, Inc., Andover, MA). ^{13}C NMR spectra were recorded at 303 K on a 10-mm broadband probe on a Bruker (Bruker Analytik GmbH, Rheinstetten, Germany) Avance 500.14 MHz. ^{13}C NMR spectra were taken with a 30-degree pulse angle, 262,144 data points, 2.5-s recovery delay, 200 scans, and 16 dummy scans.

Chemometrics. Statistical analysis of the data was done with Statistica software (Statsoft, Tulsa, OK) and included 44 analytical variables (11 FA, 14 sterols, and 19 TAG). The parameters derived from the NMR spectra formed another data set and were analyzed separately using the same approach.

The chemometric methods applied were principal component analysis (PCA), hierarchical clustering analysis (HCA), and discriminant analysis (DA), as described previously (6).

Cultivar differences were tested for each variable using a one-way ANOVA (8).

RESULTS AND DISCUSSION

Analytical determinations. The results of the analytical determinations were organized into a matrix of 37 samples × 44 variables. As a consequence of an explorative analysis of the

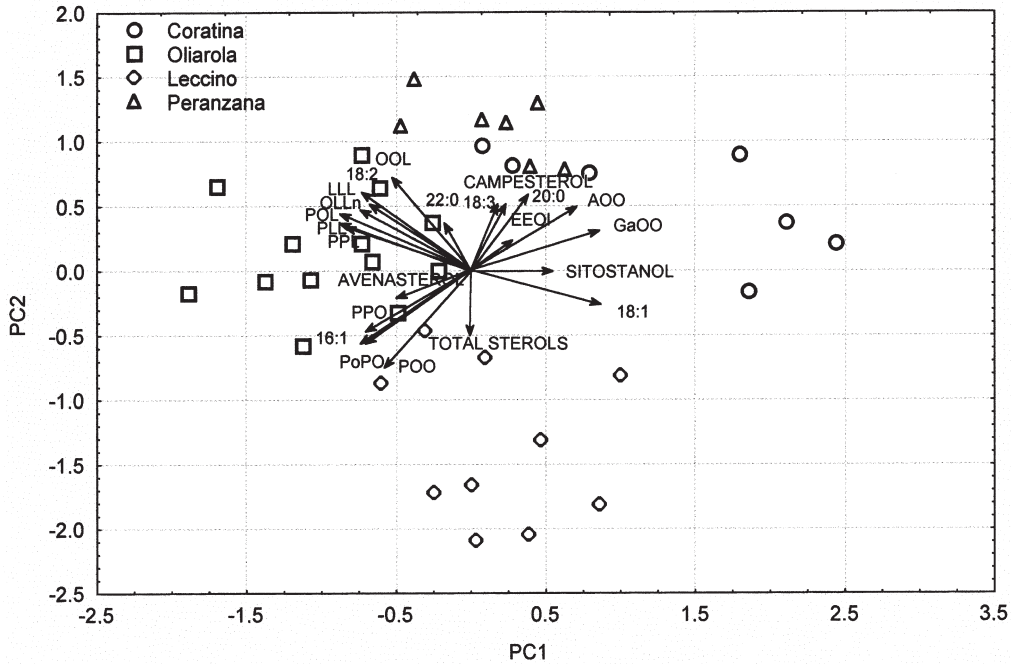

FIG. 1. Score plot and loadings from the first two principal components PC1 and PC2 obtained using FA, sterol, and TAG concentrations. L = Linoleic, O = oleic, A = arachidic, Ga = gadoleic, P = palmitic, Po = palmitoleic, Ln = linolenic.

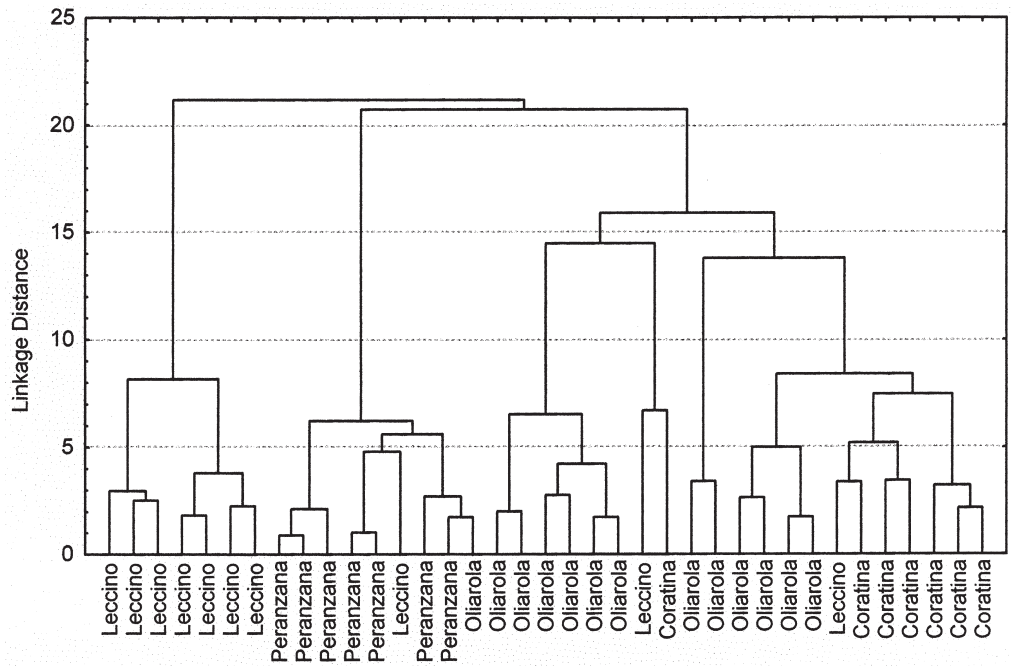

FIG. 2. Dendrogram of olive oils obtained using FA, sterol, and TAG concentrations.

data sets, the concentrations of β-sitosterol and of the TAG POO + SOL, OOO, and OLL (where P = palmitic, S = stearic, O = oleic, and L = linoleic) were eliminated because of their correlation with those of other analytes.

Since the number of the variables exceeded that of the samples, it was necessary to eliminate the parameters that did not provide relevant information for cultivar discrimination. Therefore, one-way ANOVA was applied, and 22 variables were selected. PCA was applied on the selected variables and five principal components (PC) were extracted, covering 82% of the variance.

Since a five-dimensional display presents practical difficulties, the first two PC were used for the plot in Figure 1. An examination of the loadings associated with each PC allows the most important variables to be singled out. PC1 is highly correlated with OLLn, PLL, POL, PoPO, PPL, GaOO, AOO (where Ln = linolenic, Po = palmitoleic, Ga = gadoleic, and A = arachidic), and with P, O, and L acids; OOL and POO strongly affect PC2. Samples of the Oliarola cultivar were correlated with their content of OLLn, POL, PPL, and L acid, whereas samples of Peranzana and Coratina cultivars had a high content of GaOO, AOO, O, and Ln acids. Leccino oils were more correlated with PoPO and Po acid. These variables contain information useful to distinguish the four cultivars.

The results obtained were confirmed by performing an HCA on the first five PC, applying a Ward linkage procedure on Manhattan distances. The dendrogram in Figure 2 shows that it is possible to obtain clusters composed mostly of samples of the same cultivar.

To use the classification methods adequately, the ratio between the number of variables and samples should be ≤1:3 (11). For this reason, DA was performed on the first five PC. The results of DA based on cultivar as a grouping criterion are reported in Figure 3. The predictability of the model was 95%. All the samples were correctly classified except one Coratina, which fell within the Peranzana group, and one Leccino, assigned to the Oliarola group. This confirmed the results obtained with PCA analysis.

To test the predictability of the model, samples were randomly split into a training set to develop a discriminant model and a validation set (2 from the Coratina, 4 from the Oliarola, 4 from the Leccino, and 2 from the Peranzana cultivar) on which the model could be tested. The predictability for the validation set was 92%. This result demonstrates the validity of the model.

^{13}C NMR determinations. Two approaches were applied for the quantification of signals in the ^{13}C NMR spectrum: (i) signal heights and (ii) areas of signals obtained after a Lorentzian deconvolution because of partial overlapping. In both cases, the intensities obtained were normalized to 100. The approach using areas introduced inaccuracies and had limited reproducibility. Therefore, the first approach was the only one used for the statistical analysis. The initial set of 67 signals was reduced to 12 (Table 2) after correlation and ANOVA. These chemical shifts (ppm) were assigned as follows: 172.76, C1 linoleic acid (β); 130.22, C13 linoleic acid (α); 129.87, unknown; 128.18, C10 linoleic acid (β); 34.11, C2 saturated FA; 32.00, n-3 unsaturated FA; 29.86, C12 oleic acid; 29.71, C7 saturated acid (α); 29.69, unknown; 29.58, unknown; 29.52, unknown; 14.10 and n-1 FA, where α and β indicate the 1,3- and 2-glycerol positions, respectively.

PCA analysis showed that all the variables support PC1 and PC2 (Fig. 4). In the score plot, samples of Oliarola and

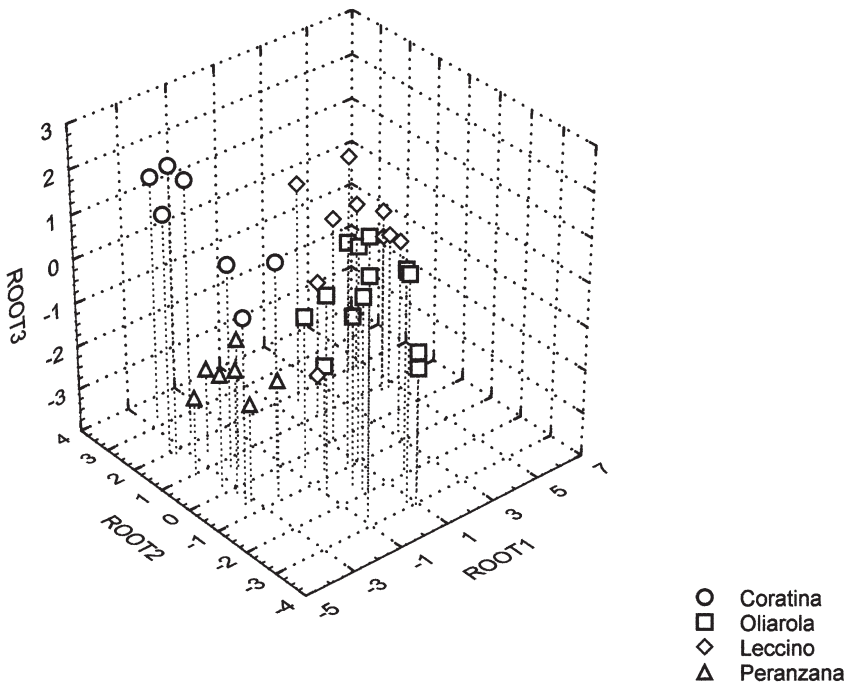

FIG. 3. Plot of the three cultivars for olive oil samples on the three discriminant functions for FA, sterol, and TAG concentrations.

Leccino cultivars are differentiated from the others, whereas there is a partial overlapping between samples of Peranzana and Coratina cultivars. Oliarola samples are characterized by positive scores on PC1, and Leccino samples are distinguished from the others because they have positive loadings on PC2. A similar differentiation was obtained with the analytical data previously discussed.

HCA, based on Euclidean distances, was applied to the first three principal components. The results of this analysis are reported as a dendrogram in Figure 5. By limiting the dendrogram

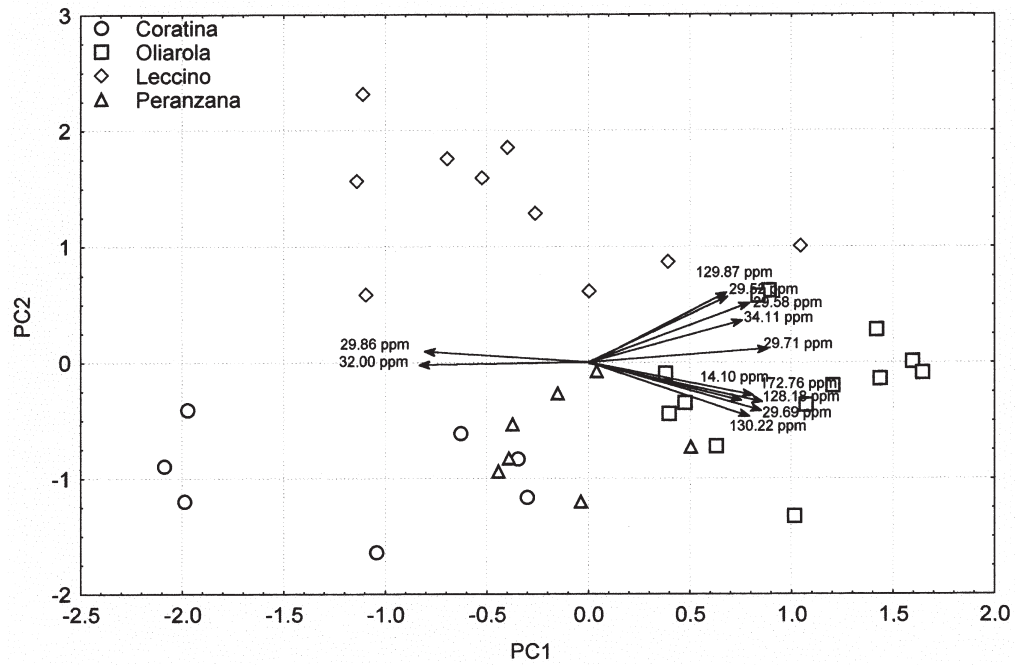

FIG. 4. Score plot and loadings from the first two principal components PC1 and PC2 obtained using ^{13}C NMR data. For abbreviation see Figure 1.

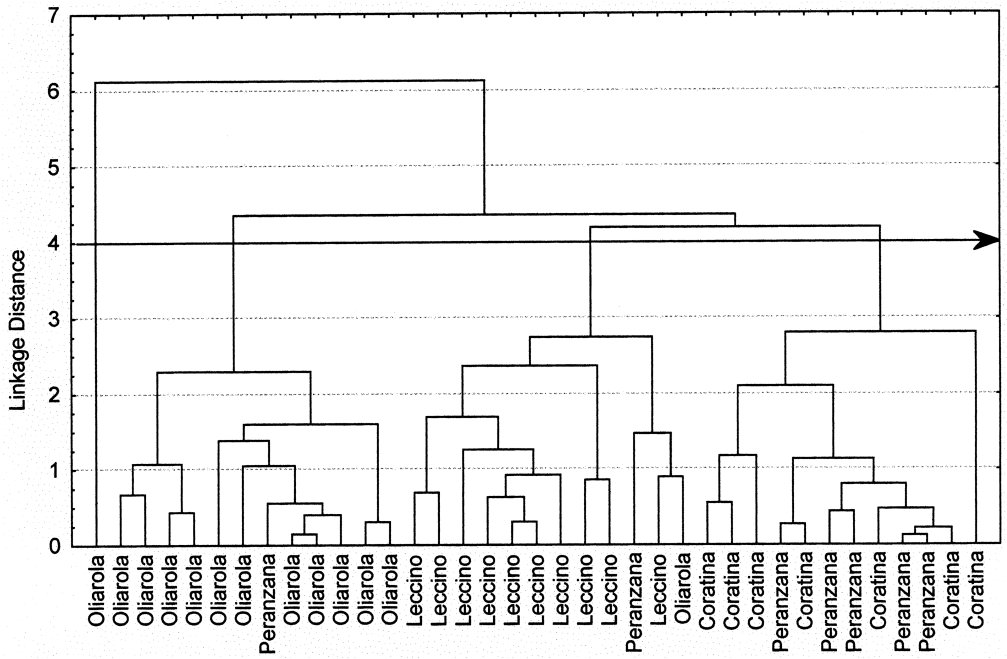

FIG. 5. Dendrogram of olive oils obtained using ^{13}C-NMR data.

at linkage distance 4, three groups were obtained. The first one consists of 11 Oliarola oils and one of the Peranzana cultivar. The second contains predominantly oils from the Leccino cultivar, and the third, oils from Peranzana and Coratina cultivars.

To evaluate the classification and predictability according to the sample cultivars, DA was applied on the data set with 25 samples in the training set and 12 in the test set. The partition of the samples in these two subsets was done randomly. The predictability of the model is *ca*. 95% (Fig. 6). Root 1 was predominantly weighted by peaks at 34.11 and 29.69 ppm, and Root 2 was marked by the intensity of the signal resonating at 128.18 ppm.

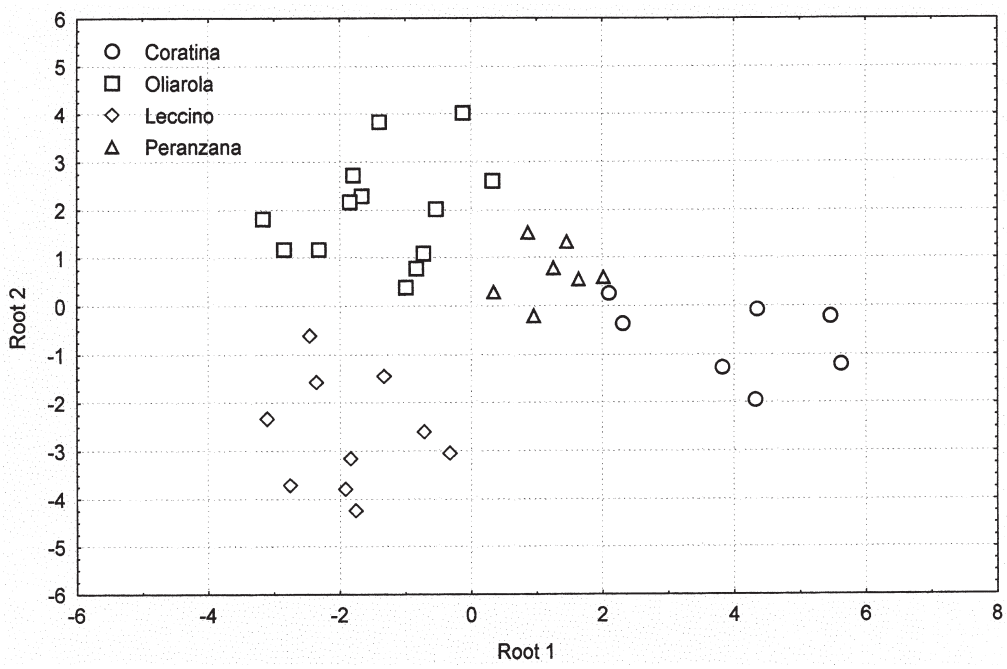

FIG. 6. Plot of the three cultivars for olive oil samples on the first two discriminant functions for ^{13}C NMR data.

The percentage of correctly predicted samples, determined with the same procedure described previously, is 100% for all cultivars except Coratina (50%).

The efficiency of official methods and ^{13}C NMR spectroscopy to similarly distinguish oil composition of olive cultivars was demonstrated. The models obtained from applying DA to the data sets of chromatographic and ^{13}C NMR results were statistically similar, and the predictability of the constructed models, determined with the test set method, were 92 and 88%, respectively. These results show that the two approaches agree. As far as the analytical data set is concerned, FA and TAG best distinguish cultivar oils. All the selected peaks in ^{13}C NMR were relevant. These data are important, because they confirm that ^{13}C NMR spectra contain the information about acyl composition and positional distribution of the glycerol moiety; therefore, this technique could be used for classifying cultivars. Nevertheless, the chromatographic data give more information about the composition, e.g., FA, of the studied oils.

ACKNOWLEDGMENTS

A.PR.OL. and ASSO.PR.OLI., Apulian oil producers associations, supplied us with some samples; Dr. Stefano Ghelli provided valuable advice; Dr. Fabiano Reniero and Dr. Claude Guillou allowed us to use the 500 MHz NMR facility of the Joint Research Centre, Institute for Health and Consumer Protection, Food Products and Consumer Goods Unit (Ispra).

REFERENCES

1. ISMEA, Filiera "Olio d'oliva," Agrisole-II Sole 24 Ore, Roma (2000).
2. Vasquez Roncero, A., Les Poliphénoles de l'Huile et leur Influence sur les Caractéristiques de l'Huile, *Rev. Franç. Corps Gras 25*:21–26 (1978).
3. Armanino, C., R. Leardi, and S. Lanteri, Chemometric Analysis of Tuscan Olive Oils, *Chemom. Intell. Lab. Sys. 5*:343–354 (1989).
4. Tsimidou, M., and K.X. Karakostas, Geographical Characterization of Greek Virgin Olive Oil by Non-parametric Multivariate Evaluation of Fatty Acid Composition, *J. Sci. Food Agric. 62*:253–257 (1993).
5. Lanza C.M., C. Russo, and F. Tomaselli, Relationship Between Geographical Origin and Fatty Acid Composition of Extra-Virgin Olive Oils Produced in Three Areas of Eastern Sicily, *Ital. J. Food Sci. 4*:359–366 (1998).
6. Sacco A., M.A. Brescia, V. Liuzzi, F. Reniero, C. Guillou, S. Ghelli, and P. Van Der Meer, Characterization of Italian Olive Oils Based on Analytical and Nuclear Magnetic Resonance Determinations, *J. Am. Oil Chem. Soc. 77*:619–625 (2000).
7. Gigliotti, C., A. Daghetta, and A. Sidoli, Indagine Conoscitiva sul Contenuto Trigliceridico di Oli Extra Vergini di Oliva di Varia Provenienza, *Riv. Ital. Sostanze Grasse 70*:483–489 (1993).
8. Shaw, A.D., A. di Camillo, G. Vlahov, A. Jones, G. Bianchi, J. Rowland, and D.B. Kell, Discrimination of the Variety and Region of Origin of Extra Virgin Olive Oils Using ^{13}C-NMR and Multivariate Calibration with Variable Reduction, *Anal. Chim. Acta 348*:357–374 (1997).
9. Holland, M.V., A. Bernreuther, and F. Reniero, The Use of Amino Acids as a Fingerprint for the Monitoring of European Wines, in *Magnetic Resonance in Food Science*, edited by P.S. Belton, I. Delgadillo, A.M. Gil, and G.A. Webb, Royal Society of Chemistry, Cambridge, 1995, pp. 136–145.
10. European Communities. Regulation 2568/91, *Off. J. Eur. Communities*, L 248 (1991).
11. Defernez, M., and E.K. Kemsley, The Use and Misuse of Chemometrics for Treating Classification Problems, *Trends Anal. Chem. 16*:216–221 (1997).

[Received March 9, 2002; accepted July 25, 2003]