Comparative Study on the Triglyceride Composition of Almond Kernel Oil. A New Basis for Cultivar Chemometric Characterization

M. L. Martín-Carratalá, C. Llorens-Jordá, V. Berenguer-Navarro, and N. Grané-Teruel*

Department of Analytical Chemistry, University of Alicante, P.O. Box 99, 03080 Alicante, Spain

Nine triglycerides (LLL, OLL, PLL, OLO, PLO, PLP, OOO, POO, and SOO; triglycerides are abbreviated using L, O, P, and S for linoleoyl, oleoyl, palmitoyl, and stearoyl fat acid radicals, respectively) in the almond kernel of 19 different cultivars have been determined by high-performance liquid chromatography. Multivariate techniques have been applied to the data from 114 chromatographic determinations. Principal component analysis efficiently reduces the number of variables so that the first two principal components explain 84.4% of the total variance. The classification obtained by the application of cluster analysis to triglyceride composition differentiates the American cultivar Texas from the rest of the cultivars. The Italian cultivars are grouped. The cultivars Achaak, Del Cid, Malagueña, Desmayo Largueta, and Chellaston form another group, and the largest group includes most of the Spanish cultivars. Discriminant analysis provides convenient functions to describe the four groups previously established by cluster analysis. The calculated classification functions correctly assign samples from the testing set to their respective groups.

Keywords: *Chemometrics; almond; triglycerides; characterization; high-performance liquid chromatography; multivariate techniques*

INTRODUCTION

The typification of almond cultivars is of great interest, for both the genetic improvement of species and their commercialization. This information would also be very useful for the industry as a tool for adulteration detection. Due to the high levels of fat in almonds (Gutfinger et al., 1972; Bhati et al., 1986), triglyceride content seems to offer some possibilities for their characterization.

Data available on triglyceride composition in almond oil are scarce (Letter, 1993), and no reference is available on the statistical multivariate analysis of this fraction. Characterization of the same set of 19 almond cultivars was carried out on the basis of their fatty acid profiles (García et al., 1996; Martín Carratalá et al., 1998) and their free amino acids (Seron et al., 1998).

The aim of the present study is to achieve this characterization on the basis of their triglyceride composition using multivariate techniques and in this way to find keys for the classification of almond cultivars into groups of similar features.

EXPERIMENTAL PROCEDURES

Samples. The set studied consists of 19 almond cultivars: 7 Spanish [Malagueña (MA), Peraleja (PE), Atocha (AT), Del Cid (DC), Desmayo Largueta (DL), Ramillete (RA), and Marcona (MR)], 3 Italian [Genco (GE), Tuono (TU), and Cristomorto (CR)], 1 Australian [Chellaston (CH)], 4 American [Texas (TE), Non Pareil (NP), Titan (TI), and Wawona (WA)], 1 Tunisian [Achaak (AC)], 1 from a Caucasian region [Primorskyi (PR)], 1 French [Ferragnes (FE)], and a hybrid (CE) obtained at CEBAS (Centro de Edafología y Biología Aplicada del Segura, Murcia, Spain). They were all cultivated and harvested under the same conditions in 1995 in the experimental fields of CEBAS. Three samples were taken from every almond cultivar, and two replicates were made on each one.

Extraction and Liquid Chromatography. The extraction of the fat was carried out in a homemade extractor similar to the commercial Soxtec according to a method described by García et al. (1996). The high-performance liquid chromatography (HPLC) analysis was performed with a Waters 600E liquid Waters 410 refractive index detector, interfaced to a computer provided with Millenium software for data acquisition and processing. A Simmetry C_{18} column (Waters), 5 μm and 250 × 4.6 mm i.d., with a guard column was used to separate the triglycerides. The optimal mobile phase found consisted of acetone and acetonitrile in the ratio 65:35, v/v. The flow rate was held constant at 1.5 mL/min at a controlled column temperature of 30 °C throughout the separation. The detector temperature was set at 40 °C. The procedure described is based on the method recommended by the European Community (EC) for the analysis of triglycerides in vegetable oils, following IUPAC (1991) method 3324.

Triglyceride peaks were identified by comparison with known oils and triglyceride standards. Olive oil, soybean oil, and a reference mixture containing four triglyceride standards were used to identify individual peaks. Triglyceride standards LLL, OLL, OOO, and SOO were purchased from Sigma.

Statistical Analysis and Data Processing. Several statistical methods have been applied to chromatographic data using an SPSS statistical package (SPSS, 1994). First, one-way ANOVA (using the Tukey-B procedure) and the Kruskal−Wallis test were applied to test the triglycerides that could contribute most to differentiate almond cultivars. The significance was estimated comparing $F_{observed}$ with F_{cited} for $(g - 1)$ and $(N - g)$ degrees of freedom, where $g = 19$ is the number of groups and $N = 114$ is the total number of determinations. Starting from this information, we sequentially applied principal component, cluster, and linear discriminant analyses.

Principal component analysis (PCA) was applied to autoscaled correlations using different criteria: scree test (Cattel, 1996) and mean eigenvalues (Cela, 1994). The cluster analysis was carried out by assuming the square of Euclidean distance as the criterion of proximity and the average linkage method for the agglomeration (Afifi and Clark, 1990). Linear discrimi-

* Corresponding author (e-mail Nuria.Grane@ua.es).

10.1021/jf981220n CCC: $18.00
Published on Web 08/07/1999

Table 1. Mean Percentage of Individual Triglyceride in the Triglyceride Fraction of Almond Oil and the Corresponding Standard Deviations

sample		LLL	OLL	PLL	OLO	PLO	PLP	OOO	POO	SOO
					Training Set					
Achaak (AC)	mean	1.96	11.34	1.73	25.68	9.54	0.44	34.27	12.61	2.43
	SD	0.02	0.09	0.02	0.16	0.09	0.01	0.17	0.14	0.05
Atocha (AT)	mean	1.53	7.97	1.44	18.81	6.52	0.40	48.43	12.09	2.79
	SD	0.02	0.08	0.01	0.05	0.05	0.01	0.10	0.03	0.01
Cebas (CE)	mean	1.58	9.92	1.45	22.66	7.64	0.39	42.16	11.43	2.77
	SD	0.01	0.05	0.02	0.09	0.05	0.01	0.12	0.07	0.03
Chellaston (CH)	mean	2.24	12.68	2.08	26.25	10.69	0.49	30.29	12.28	2.99
	SD	0.03	0.08	0.02	0.16	0.05	0.01	0.21	0.07	0.05
Cristomorto (CR)	mean	0.96	6.33	0.98	16.40	5.12	0.36	54.65	11.59	3.60
	SD	0.03	0.03	0.01	0.16	0.07	0.01	0.15	0.13	0.04
Del Cid (DC)	mean	2.12	12.29	2.00	25.37	9.58	0.49	33.39	11.90	2.85
	SD	0.01	0.05	0.01	0.12	0.07	0.01	0.13	0.04	0.04
Desmayo Largueta (DL)	mean	2.29	13.99	2.03	25.78	9.80	0.45	31.97	11.32	2.21
	SD	0.01	0.05	0.03	0.30	0.18	0.01	0.21	0.13	0.07
Ferragnes (FE)	mean	1.78	9.46	1.53	22.99	7.27	0.40	41.68	11.35	3.53
	SD	0.02	0.05	0.02	0.13	0.06	0.01	0.15	0.11	0.04
Genco (GE)	mean	1.49	6.96	1.09	17.81	5.19	0.33	52.42	11.87	2.80
	SD	0.01	0.05	0.01	0.10	0.03	0.01	0.15	0.06	0.05
Malagueña (MA)	mean	2.19	11.83	2.14	21.23	8.52	0.52	38.00	11.79	3.77
	SD	0.02	0.12	0.03	0.09	0.10	0.01	0.17	0.14	0.04
Marcona (MR)	mean	1.59	10.87	1.48	24.15	8.18	0.39	38.76	11.68	2.89
	SD	0.02	0.10	0.02	0.14	0.05	0.01	0.21	0.09	0.04
Non Pareil (NP)	mean	1.93	10.11	1.52	24.27	7.78	0.40	39.69	11.30	2.98
	SD	0.01	0.04	0.02	0.15	0.05	0.01	0.07	0.09	0.04
Peraleja (PE)	mean	2.02	10.01	1.73	20.48	7.60	0.43	42.74	12.06	2.91
	SD	0.01	0.06	0.02	0.08	0.04	0.01	0.15	0.09	0.12
Primorskyi (PR)	mean	2.54	12.56	1.81	24.50	8.54	0.38	36.65	10.75	2.28
	SD	0.01	0.11	0.02	0.12	0.04	0.01	0.08	0.08	0.04
Ramillete (RA)	mean	1.37	9.71	1.51	22.71	7.97	0.50	41.61	12.30	2.34
	SD	0.03	0.06	0.03	0.12	0.02	0.01	0.12	0.06	0.08
Texas (TE)	mean	3.63	16.93	2.79	27.49	10.95	0.43	24.70	10.31	2.76
	SD	0.03	0.06	0.02	0.16	0.06	0.01	0.11	0.09	0.04
Titan (TI)	mean	2.11	11.28	1.64	24.95	7.97	0.39	38.97	10.81	1.88
	SD	0.03	0.09	0.05	0.08	0.04	0.01	0.17	0.04	0.07
Tuono (TU)	mean	1.18	6.48	1.17	18.95	5.79	0.36	49.10	12.96	4.00
	SD	0.02	0.03	0.01	0.04	0.02	0.01	0.06	0.07	0.03
Wawona (WA)	mean	2.35	8.37	1.35	19.85	6.40	0.35	47.47	11.34	2.52
	SD	0.02	0.04	0.02	0.14	0.09	0.01	0.14	0.06	0.04
					Testing Set					
Cristomorto (CR)	mean	1.52	8.01	1.30	17.63	5.57	0.36	51.75	10.97	2.88
	SD	0.01	0.02	0.01	0.04	0.02	0.01	0.01	0.04	0.02
Desmayo Largueta (DL)	mean	2.04	11.65	1.58	26.17	9.32	0.38	35.29	11.65	1.92
	SD	0.01	0.05	0.02	0.06	0.02	0.01	0.06	0.05	0.03
Non Pareil (NP)	mean	1.92	10.10	1.50	24.18	7.78	0.39	39.68	11.32	1.94
	SD	0.02	0.01	0.02	0.10	0.03	0.01	0.04	0.04	0.08
Texas (TE)	mean	3.59	16.88	2.78	27.42	10.91	0.43	24.81	10.35	2.82
	SD	0.03	0.07	0.01	0.08	0.03	0.01	0.11	0.02	0.05

nant analisis (LDA) was conducted stepwise using the Wilks' lambda statistic (Tabachnick and Fidell, 1992) for variable selection.

RESULTS AND DISCUSSION

Triglyceride peaks were well resolved and eluted in under 30 min. The injection of oils, the composition of which was well-known (olive and soybean oil), allowed us to identify the great majority of the almond seed oil triglycerides on the basis of their retention times. Several of the peak assignments were verified with triglyceride standards LLL, OLL, OOO, and SOO. Reproducibility of analytical results ranged between 0.2 and 1.8% within a day and was better than 3% between days.

The mean values of individual triglyceride percentages in almond seed oil (Table 1) were calculated from normalized areas in the chromatograms. The triglycerides in all of the cultivars studied decrease in the following order: OOO > OLO > POO >OLL > PLO > SOO > LLL > PLL > PLP. The vast majority of triglycerides are OOO and OLO, which altogether account for >60% of the total triglycerides.

First, an exploratory study was carried out for each triglyceride individually to test whether there were significant differences between the percentages found. A nonparametric method (Kruskal−Wallis test) was used because of the small number of samples of each cultivar. The significance level obtained for the chi-squared value calculated for each triglyceride is >95% in all cases. The results have been compared with those obtained by applying one-way ANOVA (parametric method) to the data. Table 2 shows the comparison between both methods and the percentage of significant differences found when samples are compared pairwise. It can be seen that no single triglyceride allows us to differentiate among all of the cultivars. The results obtained using both methods are very similar, showing the robustness of the one-way ANOVA method even when it is applied to samples of small size.

PCA has been applied to the data previously autoscaled to reduce the number of variables and visualize

Table 2. Comparison of the ANOVA and Kruskal−Wallis Test and Percentage of Significant Differences ($p = 0.05$) When Samples Are Compared Pairwise[a]

	ANOVA		Kruskal−Wallis test	
triglyceride	*F* ratio	sig diff, %	chi-squared	sig diff, %
LLL	5190	98.2	113	98.8
OLL	8798	98.2	112	98.8
OLO	4119	98.2	112	98.8
OOO	16104	98.8	113	98.8
PLL	2104	94.7	112	94.2
PLO	3459	96.5	112	96.5
PLP	353	87.7	110	88.3
POO	313	89.5	110	88.3
SOO	633	87.1	110	87.1

[a] The significance of *F*-ratio and chi-squared values isare <lower than 0.0001.

Table 3. Loadings of the First Two Principal Components

	principal component			principal component	
variable	PC1	PC2	variable	PC1	PC2
LLL	0.141	−0.198	PLO	0.158	0.142
OLL	0.162	−0.028	PLP	0.100	0.461
OLO	0.151	0.004	POO	−0.074	0.522
OOO	−0.161	−0.083	SOO	−0.071	0.330
PLL	0.155	0.077			

the samples studied in a dimensionally more reduced space.

The correlation matrix has been calculated. As a general rule, the correlation values between triglycerides are high, except for the triglycerides POO and SOO. Triglyceride OOO shows high and negative correlation values except for POO and SOO, not surprisingly because OOO represents ∼40% of the triglycerides, and therefore it is inversely related to the rest of them. Because the determinant of the correlation matrix is $<10^{-7}$ and the Bartlett sphericity test is statistically significant ($p < 0.05$), the data of the matrix are suitable for the application of PCA.

Nine new variables are obtained when PCA is applied to the data. The first two principal components, selected according to the Kaiser criterion (eigenvalues > 1) and the sedimentation graphic (scree test), explain 67.7 and 16.7% of the variance, respectively (84.4% in all).

The communality values obtained by applying the proposed model to the data are >80% for all of the triglycerides except SOO, for which communality amounts to 43%. By application of the model, the reproduced correlation matrix has been calculated as well as the residuals between the observed and the reproduced correlations. There are only 13 residuals (36%) with absolute values >0.05; therefore, it is considered that the proposed model is adequately adjusted to the data.

The correlation values between the original variables and the principal components retained are listed in Table 3. Triglycerides OLL, OLO, OOO, PLL, and PLO show higher correlations with principal component 1 (PC1), whereas triglycerides PLP, POO, and SOO (and to a lesser extent LLL) have higher correlations with principal component 2 (PC2). If the correlations with absolute values >0.3 are considered, only triglycerides PLP, POO, and SOO are correlated with PC2. Figure 1 shows the projection of the different almond cultivars in the reduced space determined by the first two principal components. It can be seen that the American cultivar Texas and the Italian cultivars Cristomorto, Tuono, and Genco show a clearly differentiated score in the first principal component. In contrast, the rest of the cultivars differ from one another with respect to the second principal component.

Cluster analysis has been applied to classify the cultivars studied into groups. Similar results have been obtained by applying the algorithm both to the individual data and to the mean values of the different almond cultivars. Figure 2 shows the dendrogram obtained using the mean values. Four groups are obtained, at rescaled distance 7 (28%). The largest group consists of most of the Spanish cultivars (Cebas, Marcona, Atocha, Peraleja, and Ramillete) together with the French Ferragnes, the Caucasian Primorskyi, and the

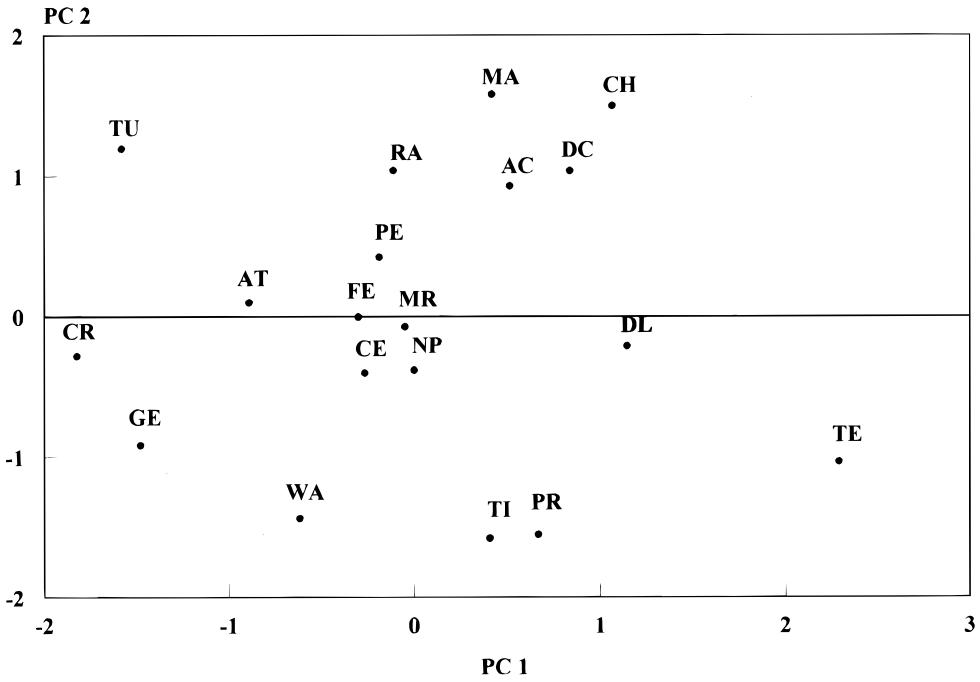

Figure 1. Plot of the average scores of almond cultivars projected on the reduced space of the first two principal components (PC1 and PC2).

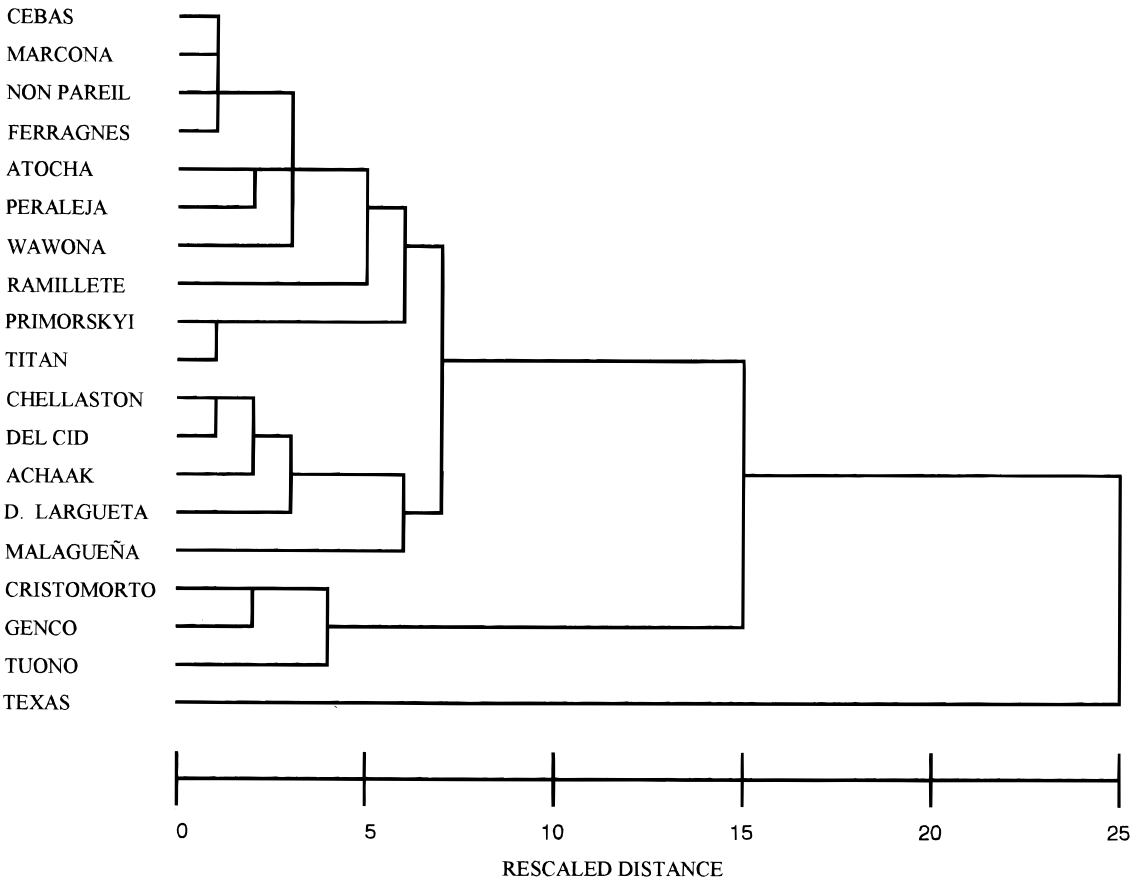

Figure 2. Dendrogram using average linkage cluster analysis.

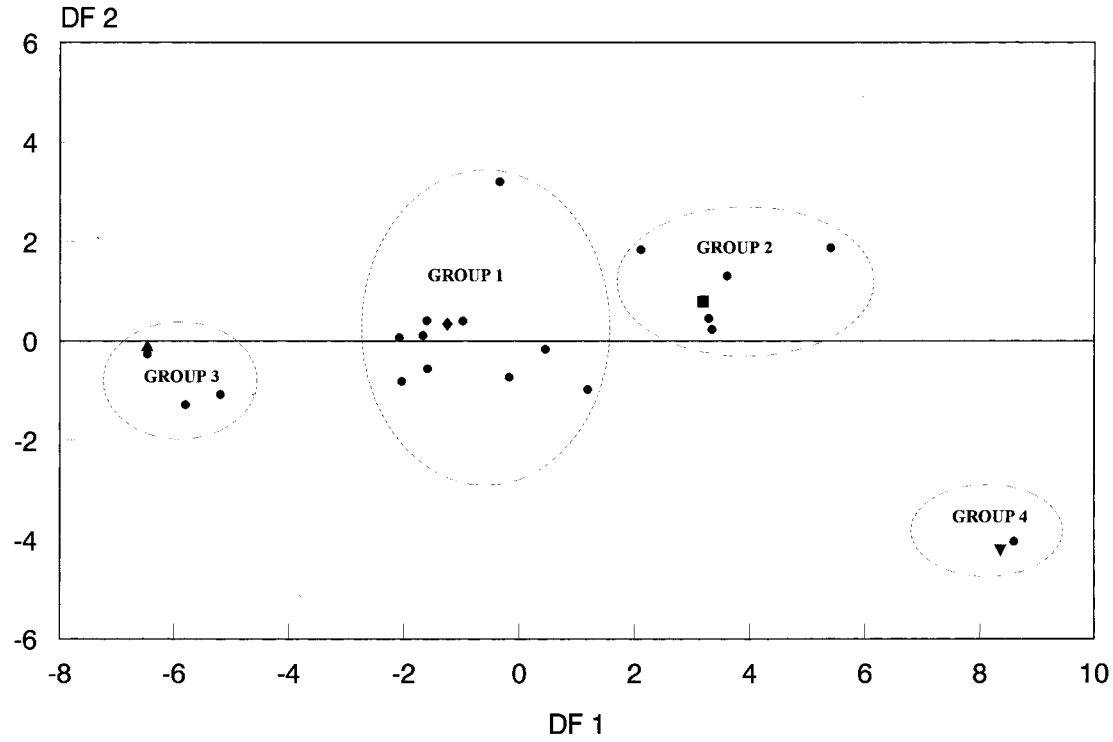

Figure 3. Plot of the mean scores of almond cultivars projected on the reduced space of the two discriminant functions (DF1 and DF2): training set (●) group 1, CE, MR, NP, FE, AT, PE, WA, RA, PR, TI; group 2, CH, DC, AC, DL, MA; group 3, CR, GE, TU; group 4, TE; testing set (▼) CR, (◆) NP, (■) DL, (▲) TE.

American cultivars Non Pareil, Wawona, and Titan. Another group includes Chellaston, Del Cid, Achaak, Desmayo Largueta, and Malagueña. The Italian cultivars Cristomorto, Genco, and Tuono are associated. The American cultivar Texas can be considered as one individual class totally separated from the rest of the cultivars. This classification agrees with the visualization previously obtained when PCA was applied.

Table 4. Eigenvalues, Variance, and Canonical Correlation Values of the Discriminant Functions

discriminant function	eigenvalue	percentage of variance	cumulative percentage	canonical correlation
1	13.44	86.44	86.44	0.96
2	1.38	8.85	95.29	0.76
3	0.73	4.71	100.00	0.65

Finally, LDA has been applied to the data, grouping them on the basis of the four associations obtained by means of cluster analysis. The Wilks' lambda and univariate *F*-ratio values for 3 and 110 degrees of freedom have been calculated for each triglyceride. All of the triglycerides show statistically significant differences between the mean values of the established groups.

Three discriminant functions have been obtained by applying stepwise discrimant analysis, using as the variable selection rule for minimizing Wilks' lambda tolerance levels of 0.001 and 1 for *F* to enter and 1 for *F* to remove. The variance explained by each discriminant function, the eigenvalues, and the canonical correlation values are presented in Table 4. The first two discriminant functions explain >95% of the total variance. Triglyceride PLO has not been included in the calculations because its tolerance level is below the established minimum. This fact indicates that this variable can be considered as a linear combination of other independent variables.

On the basis of the scores obtained for each sample as to the first two discriminant functions, the corresponding classification functions have been calculated; all of the samples studied are correctly classified into their groups. To validate the proposed model a testing set was checked, consisting of samples of the Non Pareil, Desmayo Largueta, Cristomorto, and Texas cultivars not included in the calculation of the discriminant functions. The classification obtained for them was totally correct.

Figure 3 shows the mean discriminant scores calculated for both the sample set from which the functions were calculated (training set) and the testing set.

It is apparent that in general the triglyceride composition of almond oil is very similar among different cultivars. LDA succeeds in classifying them into four different groups. Surely, groups 1 and 2 are not far from one another and both include the majority of the cultivars studied. The highlights of the multivariate analysis, however, are the singular triglyceride composition of Texas almond, on the one hand, and that of the group Cristomorto, Tuono, and Genco, on the other hand, which could be useful for discrimination purposes. We do not ignore the potential meaning of such groupings. For instance, the fact that the cultivars (GE, TU, and CR) having a known common geographic origin appear close to one another and that NP is close to Spanish cultivars and distant from TE probably suggests a different origin or an important evolution stage of these two American cultivars.

LITERATURE CITED

Afifi, A.A.; Clark, V. *Computer-Aided Multivariate Analysis*, 2nd ed.; New York, 1990; pp 429−462.

Bhati, A.; Benbouzid, M.; Hamilton, R. J.; Sewell, P. A. Separation of triglycerides and hydrocarbons from seed oils by high performance liquid chromatography with an infrared detector. *Chem. Ind.* **1986**, Jan 20, 70−71.

Cattel, R. B. The scree test for the number of factors. *Mult. Behav. Res.* **1966**, *1*, 245−276.

Cela, R. *Avances en Quimiometría Práctica*; Servicio de Publicaciones e Intercambio Científico: Santiago de Compostela, Spain, 1994; pp 241−353.

García-López, C.; Grané-Teruel, N.; Berenguer-Navarro, V.; Martín-Carratalá, M. L. Major fatty acid composition of 19 almond cultivars of different origins. A chemometric approach. *J. Agric. Food Chem.* **1996**, *44*, 1751−1755.

Gutfinger, T.; Romano, S.; Letan, A. Characterization of lipids from seeds of the rosacea family. *J. Food Sci.* **1972**, *37*, 938−939.

IUPAC. *Determination of Triglycerides in Liquid Vegetable Oils in Terms of Their Equivalent Carbon Number by High-Performance Liquid Chromatography*; International Union of Pure and Applied Chemistry: Oxford, U.K., 1991; Method 3324.

Letter, W. S. A qualitative method for triglyceride analysis by HPLC using an evaporative light-scattering detector. *J. Liq. Chromatogr.* **1993**, *16* (1), 225−239.

Martín Carratalá, M. L.; García López, C.; Berenguer Navarro, V.; Grané Teruel, N. New contribution to the chemometric characterization of almond cultivars on the basis of their fatty acid profiles. *J. Agric. Food Chem.* **1998**, *46*, 963−967.

Seron, L. H.; Garrigós, E.; Prats, S.; Martín, M. L.; Berenguer, V.; Grané, N. Chemometric characterization of almond cultivars on the basis of their free-amino acid profiles. *Food Chem.* **1998**, *61* (4), 455−459.

SPSS. Release 6.0.1.; SPSS Inc., Chicago, IL, 1994.

Tabachnick, B.; Fidell, L. *Using Multivariate Statistics*, 12th ed.; Harper Collins Publishers: New York, 1992; pp 505−596.

Received for review November 10, 1998. Revised manuscript received June 10, 1999. Accepted June 14, 1999.

JF981220N