A Possible Way to Predict the Genetic Relatedness of Selected Almond Cultivars

N. Grane-Teruel, M.S. Prats-Moya*, V. Berenguer-Navarro, and M.L. Martin-Carratala

Department of Analytical Chemistry, University of Alicante, 03080 Alicante, Spain

ABSTRACT: Some multivariate statistical techniques were applied to the fatty acid composition of almond seed for a set of 107 samples of cultivars grown in different environmental conditions. These techniques reveal a close similarity in fatty acid composition among a cultivar and its ancestors as in the case of Ferragnes and Masbovera. Interestingly, the Guara cultivar seems to be very similar to the Tuono cultivar. Another finding suggests that the typical Spanish cultivars Desmayo Largueta and Marcona have a similar fatty acid profile to the American cultivar Non Pareil. Independent of environmental conditions, this shows a possible common parentage in the pedigree of these cultivars.

Paper no. J9834 in *JAOCS 78*, 617–619 (June 2001).

KEY WORDS: Almond, chemometrics, fatty acids, gas chromatography.

Almonds are known oleaginous seeds and are very appreciated for their nutritional and sensory characteristics. Almond fat is also used in valuable cosmetic preparations. For many years agricultural research programs around the world have introduced new cultivars that improve not only the production but also the quality of the seed. In Spain, three research centers (Servicio de Investigacion Agraria, SIA, in Zaragoza; Centro Agropecuario de Mas Bove in Reus; and Centro de Edafología y Biologia Aplicada del Segura, CEBAS, in Murcia) are particularly active in production of new plant materials. All three organizations have succeeded in finding new almond cultivars, such as Guara or Masbovera, that have been commercialized successfully and today grow in major areas of Spain. We have studied some chemical components of these cultivars in comparison with other traditional varieties (1). The use of statistical multivariate techniques has revealed interesting compositional relations among these cultivars. The aim of this work is to test the hypotheses that fatty acid composition is characteristic of groups of cultivars, independent of growing conditions, and that such chemical similarities may imply a genetic relationship within an extensive set of cultivars grown in different regions and environmental conditions.

MATERIALS AND METHODS

The set of samples selected in this work consisted of 10 cultivars: Desmayo Largueta (DL), Guara (GU), Marcona (MR), and Masbovera (MS) from Spain; Texas (TE), Non Pareil (NP), and Titan (TI) from the United States; Tuono (TU) from Italy; Ferragnes (FE) from France; and Primorskyi (PR) from a Caucasian region. The ancestors for the cultivars Masbovera, Ferragnes, and Primorskyi are known to be Primorskyi × Cristomorto, Ai × Cristomorto, and Princesse × Nikitskyi, respectively. The cultivars were grown in different geographical areas: Maria (Almeria, ALM), La Puebla de Don Fabrique (Granada, GR), Santomera and Cehegin (Murcia, MUR), Castalla and Bacarot (Alicante, ALI), Mas Valero, Mas Bove, and Mora d'Ebre (Tarragona, TAR), and Aula Dei (Zaragoza, ZAR), as well as Avignon (AVIG) in France. These locations differ significantly in climatological conditions. Granada and Zaragoza are cold regions, while the other localities have mild climates. The total number of samples collected was 107.

Almond samples were blanched and ground in an electric grinder. Oil was extracted in a home-made modified Soxhlet apparatus, using a mixture of diethyl ether/hexane (1:1 vol/vol). Afterward, the fatty acids contained in 2 g of almond oil were transformed into their respective fatty acid methyl esters. The methylation of the lipids was done according to the American Oil Chemists' Society method (2).

Experimental data were obtained by using a Carlo Erba Series 8000 gas chromatograph equipped with a split/splitless injector and a flame-ionization detector (FID) and interfaced to a computer provided with a program for data acquisition and processing (Chrom-card; Fisons Instruments, Poole, United Kingdom). The chromatographic column was a cyanopropyl methyl siloxano (Quadrex Corporation, New Haven, CT), 30 m × 0.25 mm i.d. Samples of 1 μL were injected into the split injector at a 1:30 ratio and 230°C. The carrier gas was helium at a flow rate of 2 mL min^{-1}. The FID detector was set at 300°C. Oven temperature was maintained at 110°C for 2 min, ramped from 110 to 180ºC at 10°C min^{-1}, then held at 180ºC for 2 min, ramped again from 180 to 220ºC and finally held at 220ºC for 5 min. C13:0 was used as internal standard. Identification of fatty acids in the samples was achieved by comparison with relative retention times in a reference sample that contained standard methyl esters obtained from Sigma Chemical Co. (St. Louis, MO).

Experimental data were processed with the aid of the SPSS statistical package (3). Principal component analysis was applied to autoscaled data. The components retained were selected using the Scree test and Kaiser criteria. Cluster analysis was carried out by applying the Ward method for agglom-

*To whom correspondence should be addressed at Analytical Chemistry Department, University of Alicante, P.O. Box 99, 03080 Alicante, Spain.
E-mail: maria.prats@ua.es

eration with the square of Euclidean distance as the criterion of proximity (4). A linear discriminant analysis was conducted stepwise by employing the Wilks' lambda statistics (5) for variable selection. In all cases, the algorithms used were applied to the mean values obtained from four replications for each sample.

RESULTS AND DISCUSSION

Twelve fatty acids were identified in the almond samples. These fatty acids in decreasing concentration level in the almond are: C18:1, C18:2, C16:0, C18:0, C16:1, C17:1, C17:0, C14:0, C20:0 + C18:3, C15:0, and C22:0. In general, C18:1 represented more than 60% of the total fatty acid content in all samples. C18:2, C16:0, and C18:0 plus C16:1 typically accounted for 39% of the oil. Fatty acid mean values along with their relative standard deviations for the 10 almond cultivars considered are shown in Table 1. C20:0 and C18:3 were not totally separated during gas chromatography.

A principal component analysis was applied to visualize similarities and differences among these 10 almond cultivars. The eigenvalues obtained from the correlation matrix were: 3.73, 2.89, 1.07, 1.01, 0.71, 0.69, 0.41, 0.21, 0.16, 0.12, and 0.01. After application of the Kaiser criteria and the Scree test, only four principal components were retained, accounting for 79% of the total variability. The percentage of variance for the four principal components (PC) was 33.9% for the first PC, 26.3% for the second PC, 9.7% for the third PC, and 9.2% for the last PC.

When the scores of the samples for the first two PC were represented, it was possible to visualize a clear relationship between the samples of the cultivars Guara and Tuono, and also a peculiar behavior associated with the samples belonging to the cultivars Masbovera and Ferragnes.

The next step was to apply cluster analysis to obtain the association between samples on the basis of nearness criteria among objects. A summary of the resultant dendrogram (Fig. 1) shows three groups at a rescaled distance of 10. The first group contained all samples of the cultivars Guara and Tuono, plus three samples of Desmayo Largueta, three samples of Marcona, three samples of Ferragnes, and one of Texas. The second group was formed by all samples of the cultivar Masbovera together with 12 samples of Ferragnes, 3 of Texas, 2 of Primorskyi, and 3 of Marcona. The third group contained all samples of the cultivars Non Pareil and Titan, and 13 samples of Desmayo Largueta, 11 of Marcona, 1 of Ferragnes, 2 of Primorskyi, and 4 of Texas.

Based on this information, a linear discriminant analysis was conducted according to a stepwise method with the following three groups: Guara and Tuono (group 1); Masbovera and Ferragnes (group 2); and Desmayo Largueta, Marcona, and Non Pareil (group 3). Texas, Titan, and Primorskyi samples were not included in the calculation of the discrimination functions because, after cluster analysis, their samples were not clearly classified in a group and, in the case of the cultivar Titan, because of the small number of samples (four) that we obtained.

Two discriminant functions were obtained using the variable selection rule for minimizing Wilks' lambda (a tolerance level of 0.001 for F to enter higher than 1.000, and F to remove lower than 0.999). The variance explained by each discriminant function, the eigenvalues, and the canonical correlation values are presented in Table 2. C17:0, C17:1, C18:0, C18:2, and C22:0 were not included in the calculations because their tolerance level was below the established minimum.

TABLE 1
Fatty Acid Mean Values (g/100 g sample) and Relative Standard Deviations for the Cultivars Studied

Cultivar[a]		C14:0	C15:0	C16:0	C16:1	C17:0	C17:1	C18:0	C18:1	C18:2	C[b]	C22:0
DL	Mean	0.02	0.007	4.4	0.35	0.04	0.06	1.3	38.0	11.8	0.015	0.005
	%RSD[c]	26.7	19.0	13.8	39.4	19.6	11.88	31.3	9.5	26.2	20.7	31.1
FE	Mean	0.02	0.006	4.2	0.32	0.04	0.05	1.3	40.7	12.1	0.015	0.006
	%RSD	28.2	17.0	11.4	38.0	14.1	9.02	23.0	8.7	23.2	13.8	16.9
GU	Mean	0.02	0.007	4.0	0.29	0.04	0.05	1.3	38.5	10.8	0.015	0.006
	%RSD	31.2	27.1	11.0	34.4	12.0	9.15	22.4	6.7	23.9	26.3	35.5
MR	Mean	0.02	0.007	4.2	0.34	0.04	0.05	1.3	38.4	11.8	0.014	0.006
	%RSD	18.5	13.1	10.0	29.8	32.3	9.93	22.9	8.1	23.3	22.0	30.0
MS	Mean	0.02	0.006	4.0	0.39	0.04	0.05	1.3	43.9	9.6	0.014	0.006
	%RSD	11.2	7.32	9.4	12.3	5.3	12.3	12.8	2.9	2.3	21.3	15.2
NP	Mean	9.77	0.007	6.2	12.9	0.05	3.54	1.6	38.5	14.6	0.015	0.005
	%RSD	32.1	32.4	32.1	35.8	36.5	26.5	21.9	10.5	54.3	44.3	26.3
PR	Mean	0.03	0.008	4.7	0.46	0.06	0.06	1.5	41.4	10.6	0.011	0.004
	%RSD	7.72	7.65	8.1	22.2	39.45	15.0	20.9	8.6	41.8	5.8	18.3
TE	Mean	0.02	0.007	4.2	0.33	0.04	0.05	1.2	36.9	12.2	0.013	0.005
	%RSD	20.5	17.4	8.4	40.6	11.20	8.78	36.0	5.2	13.5	18.0	29.2
TI	Mean	0.02	0.007	4.1	0.36	0.04	0.05	1.3	38.2	12.2	0.012	0.005
	%RSD	10.5	13.1	9.3	46.3	18.43	10.9	31.8	4.5	13.2	4.1	32.5
TU	Mean	0.02	0.007	4.2	0.30	0.04	0.05	1.2	38.9	12.6	0.014	0.005
	%RSD	19.4	23.2	6.4	29.3	11.88	9.32	26.0	5.2	19.3	23.8	22.3

[a]Cultivars: DL, Desmayo Largueta; GU, Guara; MR, Marcona; MS, Masbovera; TE, Texas; NP, Non Pareil; TI, Titan; TU, Tuono; FE, Ferragnes; and PR, Primorskyi.
[b]C20:0 + C18:3.
[c]% RSD, percentage of relative standard deviation.

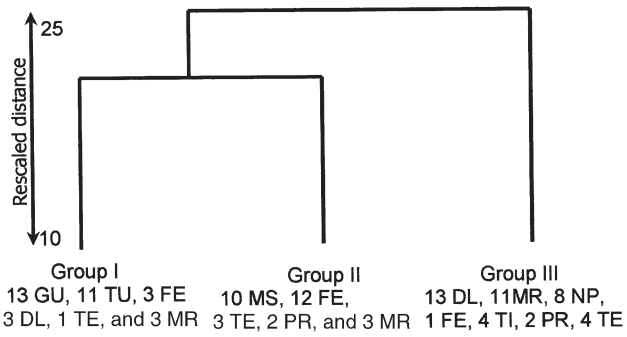

FIG. 1. Dendrogram obtained from cluster analysis. GU, Guara; TU, Tuono; Fe, Ferragnes; DL, Desmayo Largueta; TE, Texas; MR, Marcona; MS, Masbovera; PR, Primorskyi; NP, Non Pareil; TI, Titan.

By using the discriminant functions, the samples in the training set were correctly classified into their groups, except two samples of Desmayo Largueta and one of Ferragnes. Figure 2 shows the scores of almond cultivars projected on the reduced space of the two discriminant functions.

To validate the proposed model, 10 different testing sets were used to obtain a correct classification of the samples in more than 95% of the cases. The classification functions obtained were:

$$(GU + TU) = 3488\ C14{:}0 + 6711\ C15{:}0 + 135\ C16{:}0 - 121\ C16{:}1 + 23\ C18{:}1 + 2158\ (C20{:}0 + C18{:}3) - 1498 \quad [1]$$

$$(MS + FE) = 3231\ C14{:}0 + 6393\ C15{:}0 + 130\ C16{:}0 - 93\ C16{:}1 + 22\ C18{:}1 + 1856\ (C20{:}0 + C18{:}3) - 1364 \quad [2]$$

$$(DL + MR + NP) = 3446\ C14{:}0 + 6795\ C15{:}0 + 132\ C16{:}0 - 82\ C16{:}1 + 22\ C18{:}1 + 1758\ (C20{:}0+C18{:}3) - 1383 \quad [3]$$

In using these classification functions, the samples of the cultivars Texas and Primorskyi, not included in the calculation of the discrimination functions, are all classified in group 2 together with the cultivars Masbovera and Ferragnes. And the samples of the cultivar Titan are classified in group 3 together with Marcona, Desmayo Largueta, and the American Non Pareil.

From this study we were able to establish a classification of the 10 cultivars selected into three groups. One of the groups is established by the cultivars Guara and Tuono, showing that the cultivar Guara may have been derived from the cultivar Tuono. In group 2 Masbovera and Ferragnes cultivars are classified together; this can be justified because both have a common ancestor, the cultivar Cristomorto. The cultivar Primorskyi is also an ancestor of the cultivar Masbovera. The cultivar Texas, whose origin is unknown, could also have as an ancestor the cultivar Cristomorto or any of the other three cultivars. Finally, it is very interesting to observe that two typical Spanish cultivars Marcona and Desmayo Largueta show great similarity to the American cultivars Non Pareil and Titan.

TABLE 2
Variance Explained by Each Discriminant Function, the Eigenvalues, and the Canonical Correlations

Function	Eigenvalue	% Variance	Canonical correlation
1	4.24	83.96	0.9001
2	0.82	16.04	0.6702

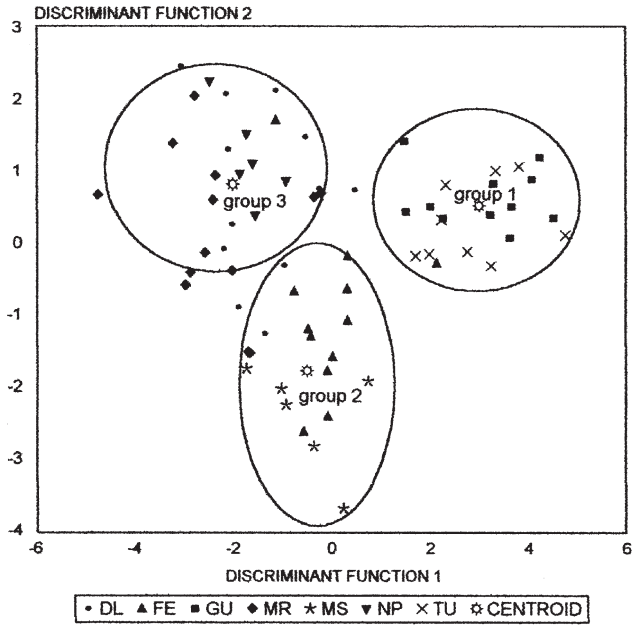

FIG. 2. Mean scores of almond cultivars for the two discriminant functions. For abbreviations see Figure 1.

REFERENCES

1. Prats-Moya, M.S., N. Grane-Teruel, V. Berenguer-Navarro, and M.L. Martín Carratala, A Chemometric Study of Genotypic Variation in Triacylglycerol Composition Among Selected Almond Cultivars. *J. Am. Oil Chem. Soc. 76*:267–272 (1999).
2. American Oil Chemists' Society, *Official and Tentative Methods of the American Oil Chemists' Society*, American Oil Chemists' Society, Chicago, 1969, Method Ce-2-66.
3. *SPSS*, release 6.0.1, SPSS Inc., Chicago, 1994.
4. Afifi, A., and V. Clark, *Computer-Aided Multivariate Analysis*, 2nd edn., Van Nostrand Reinhold, New York, 1990, pp. 429–462.
5. Tabachnick B., and L. Fidell, *Using Multivariate Statistics*, 12th edn., Harper Collins Publishers, New York, 1992, pp. 505–596.

[Received December 11, 2000; accepted February 28, 2001]