

Orange, Mandarin, and Hybrid Classification Using Multivariate Statistics Based on Carotenoid Profiles

K. L. Goodner,^{*,†} R. L. Rouseff,[‡] and H. J. Hofsommer[§]

Citrus and Subtropical Products Laboratory, Agricultural Research Service, U.S. Department of Agriculture, Winter Haven, Florida 33881; Citrus Research and Education Center, University of Florida, Lake Alfred, Florida 33850; and Gesellschaft für Lebensmittel-Forschung, Landgrafenstrasse 16, D10787 Berlin, Germany

A study was undertaken to differentiate citrus on the basis of a carotenoid profile obtained from the HPLC determination of 12 carotenoids found in saponified fresh juice. Mandarin, orange, and various hybrid varieties were analyzed to determine their carotenoid profiles. The resulting peak areas were analyzed using principal component analysis (PCA), canonical discriminate analysis (CDA), and Mahalanobis distances. These were used to develop models for classifying the juices into appropriate groups. Thirty-two samples were analyzed to determine classification techniques. Mandarin and orange juices were quite distinct, with the hybrids scattered throughout the mandarin and orange clusters using PCA. CDA resulted in three distinct groups with only a few of the hybrids in the orange grouping.

Keywords: Mandarin; orange; juice; statistical; classification; carotenoids; HPLC; saponification

INTRODUCTION

Citrus fruits are a complex source of carotenoids with the largest number of carotenoids found in any fruit (1). Approximately 115 different carotenoids have been reported in citrus, including a large number of isomers (2). Citrus carotenoids were extensively studied in the 1960s and early 1970s by numerous investigators who used open column chromatography to separate and identify these pigments (3–6). Whereas large amounts of material could be recovered for identification purposes, separation times required several days and artifact formation was a problem because carotenoids are sensitive to heat, light, and oxygen. Stewart and Wheaton (7) demonstrated that high-pressure liquid chromatography (HPLC) could be used to separate saponified citrus carotenoids in only a few hours. They employed hand-packed MgO columns to separate carotenes and hand-packed ZnCO₃ columns to separate xanthophylls. They subsequently used this technique to show that two previously reported citrus carotenoids were artifacts that were unintentionally produced during the saponification step (8). HPLC has since become the method of choice for the study of carotenoids regardless of their source.

The large number of carotenoids in citrus suggests that it might be possible to use the carotenoid profile for the purpose of classifying fruit. However, a literature search produced only one reference on classification of citrus based on carotenoid profiles. A recent study by Mouly and co-workers (9) differentiated Valencia orange juices from differing geographical regions using caro-

tenoid profiles and multivariate analyses. Their work showed good statistical separation for the juices from Belize, Israel, Spain, and the Florida/Cuba groups. The Florida and Cuba groups were not completely separated into groups. The closest example of using multivariate statistics to differentiate citrus varieties is the work of Sawamura et al. (10), who used cold-pressed oil (CPO) components from 8 varieties of pummelos and 29 other citrus samples.

The idea of classifying juices based on statistical models based on chemical composition is not new. Flavanones and methoxylated flavones have been used as a means of differentiating citrus juices, oils, and mixtures of citrus juices (11–15). Several researchers have used statistical models based on the volatile constituents of citrus juices to classify juices into various categories. Pino et al. (16) (24 samples) and Jella et al. (17) (29 samples) have used multivariate statistics to classify grapefruit juices into different quality classifications using gas chromatographic (GC) data coupled with sensory information (16, 17). This has significant applicability to quality assurance and for the improvement of consumer acceptance. Similarly, Shaw et al. (18) used GC data along with multivariate statistics to classify orange and grapefruit juices (27 samples) into various processing categories. These researchers were able to differentiate unpasteurized juice, pasteurized not from concentrate juice, and juice reconstituted from concentrate.

In this paper, the findings of applying multivariate statistics to a 12-component carotenoid profile of various orange, mandarin, and hybrid varieties of citrus will be discussed along with the various models and their effectiveness at accurate classifications. The data used for this analysis were first reported by Rouseff et al. (19), where they were separated and identified.

* Author to whom correspondence should be addressed [telephone (863) 293-4133, ext. 127; fax (863) 299-8678; e-mail goodner@citrus.usda.gov].

[†] U.S. Department of Agriculture.

[‡] University of Florida.

[§] Gesellschaft für Lebensmittel-Forschung.

Table 1. Citrus Varieties Studied along with Class and Variety Number Used in Figure 2

variety no.	variety name	class ^a
1	Nova Tangelo	H
2	Murcott	H
3	Orlando Tangelo	H
4	Ortanique	H
5	Sunshine Tangelo	H
6	Page	H
7	Minneola	H
8	Wekiwa	H
9	Lee	H
10	Osceola	H
11	Ugli	H
12	Orlando Tangerine	H
13	Temple	H
14	Umatilla Tangor	H
15	Batangus Tangerine	M
16	Dancy Tangerine	M
17	Empress Mandarin	M
18	Fairchild	M
19	King Mandarin	M
20	Fortune	M
21	Pua	O
22	Algerian Navel	O
23	Rhode Red Valencia	O
24	Midsweet	O
25	Valencia Orange	O
26	San Blood Orange	O
27	Ruby Blood Orange	O
28	Sanguinelli	O
29	Cara Cara Navel	O
30	Hamlin Orange	O
31	Sunstar	O
32	Late Navel Orange	O

^a H, hybrid; M, mandarin; O, orange.

MATERIALS AND METHODS

Reagents and Standards. All reagents used were of HPLC grade from Fisher Scientific or Merck. One mixture of standards purchased from Sigma Chemical Co. (St. Louis, MO) consisted of canthaxanthin, ethyl β -apo-8'-carotenoate, β -apo-8'-carotenal, lycopene, and β -carotene. Another set of standards consisting of β -cryptoxanthin, α -carotene, β -carotene, lutein, and zeaxanthin was obtained from Dr. Gerry Spinwall of Hoffmann-La Roche. Carotenoids that were not available were identified by their spectral characteristics reported in the literature.

Sample Preparation. Thirty-two varieties of citrus, which were obtained at commercial maturity from a varietal arboretum, were used in this study. The specific varieties are detailed in Table 1. Table 1 also has the general classes that were assigned to the samples for statistical analysis along with a number that corresponds to the numbers used in Figure 2. Twenty-five milliliters of single-strength hand-squeezed orange juice or 5.0 g of orange juice concentrate (reconstituted to single strength) was precipitated with Carrez solution [potassium cyanoferrate(II), 15% w/v aqueous, Carrez I]. Two milliliters of ZnSO₄·7H₂O (30% w/v aqueous, Carrez II) solution was added to 25 mL of juice and mixed. Two milliliters of K₄[Fe(CN)₆]·3H₂O was added with agitation. After mixing, the solution was allowed to stand for 10 min before being centrifuged. The supernatant was decanted and discarded. A small amount of acetone was added to solubilize the carotenoids in the precipitate. The acetone layer was removed and washed by shaking with petroleum ether and water in a separatory funnel. The petroleum ether (bp 40–60 °C) layer was separated and evaporated to dryness. The residue was dissolved in 6 mL of diethyl ether and 6 mL of 10% methanolic KOH. After standing overnight, protected from light at room temperature, it was extracted with 20 mL of diethyl ether. One hundred milliliters of a 10% NaCl solution was added to the separatory funnel. After shaking, the ether layer was removed and washed with distilled water until free of alkali. The ether layer was dried with sodium sulfate and evaporated to dryness

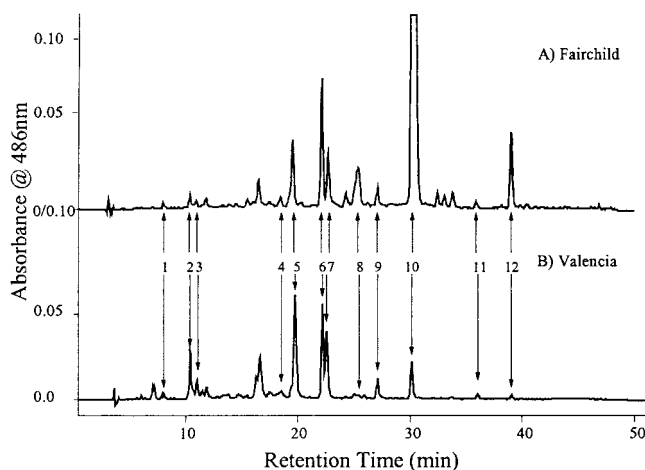


Figure 1. Two chromatograms with the 12 components used in this study: (A) Fairchild mandarin; (B) Valencia orange.

under vacuum. Carotenoids were dissolved with 0.5 mL of acetone or methyl *tert*-butyl ether (MTBE), diluted in 1.0 mL of methanol, and placed in sealed amber vials until analysis.

Equipment. The chromatographic equipment consisted primarily of a Hewlett-Packard 1090M microprocessor-controlled integrated chromatographic system consisting of three low-pressure syringe metering devices for ternary gradient formation, a single high-pressure diaphragm pump with pulse dampener, a variable-volume autosampler, and a photodiode array detector. The photodiode array detector was set to scan from 250 to 550 nm. Four separate data channels were set to record the absorbances at 290, 350, 430, and 486 nm with spectral bandwidths of 8 nm. These choices of these wavelengths are explained in detail in Rouseff et al. (19). These wavelengths were chosen as the most common peak maxima for the carotenoids of interest. Data were collected, stored, and integrated using the Hewlett-Packard HP-79994 chromatographic workstation and related software.

Chromatographic Conditions. For analytical HPLC a YMC column (4.6 mm i.d. × 25 cm, Hampsted, NC) was used, which consisted of polymeric C-30 material that was chemically bound to 5 μ m silica and not end-capped. The initial solvent composition consisted of 90% MeOH, 5% water, and 5% MTBE. The solvent composition changed in a linear fashion to 95% MeOH and 5% MTBE at 12 min. During the next 8 min (20 min of running time), the solvent composition was changed to 86% MeOH and 14% MTBE. After reaching this concentration, the solvent was gradually changed to 75% MeOH and 25% MTBE at 30 min. Final composition was reached at 50 min and consisted of 50% MeOH and 50% MTBE. All solvent changes were made in a linear fashion. Initial conditions were re-established in 2 min and re-equilibrated for 15 min before the next injection. Flow rate was 1 mL/min. Injection volume was 20 μ L.

Statistical Analyses. Statistical analyses were conducted using Statistica (Statsoft, Tulsa, OK, version 5). All statistical analyses are based on the peak areas reported from Grams/32 (Galactic Industries Corp., Salem, NH, version 4.01, level 1).

RESULTS AND DISCUSSION

HPLC Carotenoid Profiles. Figure 1 shows chromatograms of Fairchild mandarin and Valencia orange juice samples using UV absorbance at 486 nm. A wavelength of 486 nm was chosen to maximize absorbance in the red/orange region of the visible spectrum, which is the ideal area for studying carotenoids of orange, mandarin, and hybrid juice samples. Upon inspection, these chromatograms have some differences that are apparent. The most obvious difference between the two chromatograms is the peak labeled "10" in the

Table 2. Spectral Characteristics of Orange Juice Carotenoids Used for PCA

peak	carotenoid	RT ^{a,b} (min)	observed (nm)			literature (nm)			ref
			peak I ^b	peak II ^b	peak III ^b	peak I	peak II	peak III	
1	neochrome	7.97 (0.27)	399.7 (3.8)	420.0 (3.8)	446.0 (4.9)	397	420	444	3
2	antherxanthin	s0.33 (0.37)	s422.0 (3.8)	442.7 (2.9)	470.5 (2.7)	416	441	468	23
3	cis-antherxanthin	10.91 (0.39)	416.0 (3.2)	438.4 (2.5)	468.5 (2.8)	418	442	470	21
4	mutatoxanthin A	18.30 (0.63)	s406.5 (0.0)	426.6 (0.4)	451.3 (0.9)	404	427	452	21
5	lutein	19.40 (0.53)	s423.8 (3.7)	444.6 (2.6)	470.3 (2.5)	s423.5	445.5	473.5	22
6	zeaxanthin	21.96 (0.59)	s427.3 (1.1)	450.2 (0.5)	476.5 (0.2)	s425.5	451.5	477.5	22
7	isolutein	22.48 (0.67)	s418.7 (0.5)	440.5 (0.0)	466.8 (0.6)	416	440	470	24
8	unknown	25.18 (0.58)	s422.9 (1.9)	443.7 (2.0)	468.9 (2.0)				
9	α -cryptoxanthin	26.96 (0.54)	s423.0 (1.1)	444.2 (0.8)	472.3 (0.7)	420	444	472	21
10	β -cryptoxanthin	30.09 (0.60)	s428.3 (0.8)	450.5 (0.0)	476.6 (0.4)	426	452	478	23
11	α -carotene	35.79 (0.50)	s422.0 (2.8)	445.2 (0.9)	472.5 (1.2)	s423.5	445.5	473.5	22
12	β -carotene	38.91 (0.55)	s427.7 (2.7)	451.3 (1.1)	477.3 (1.1)	s429.5	451.5	477.5	22

^a RT, retention time; s, spectral shoulder. ^b Numbers in parentheses are standard deviations.

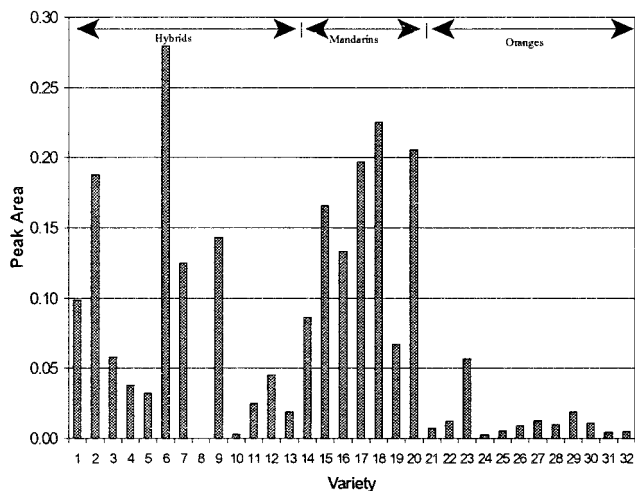


Figure 2. Bar graph showing how β -cryptoxanthin varies with respect to varietal type. Numbers correspond to the numbers in Table 1.

two chromatograms. This is β -cryptoxanthin, which has a dark orange, almost red, appearance that gives mandarin fruit their distinctive color. Some other obvious differences between these two samples are peaks 2, 8, and 12. However, casual inspection will not produce a systematic method of differentiating the juices based on these labeled peaks due to varietal variations not shown in Figure 1.

Spectral characteristics from the peaks labeled 1–12 in Figure 1 are shown in Table 2. The spectral characteristics are compared with those reported in the literature showing that the chosen peaks matched those reported with an average difference of <0.4%. These are the 12 peaks that were chosen for the statistical analyses. The β -cryptoxanthin peak mentioned previously is one of the key compounds for differentiating among varieties. The distribution of β -cryptoxanthin for mandarins, oranges, and hybrid varieties reveals a mean (median) of peak areas of 0.166 (0.182), 0.0130 (0.009), and 0.082 (0.051), respectively. The individual peak areas are shown in Figure 2. As can be seen, β -cryptoxanthin values vary drastically within the different hybrid varieties, being large for all of the mandarin juices and low for all varieties of orange juice with the exception of the Rhode Red variety. The Rhode Red variety is a naturally occurring mutation that is named after its discoverer Paul Rhode, Sr., and has characteristics similar to those of the Valencia in appearance, flavor, and time of maturity, but with a much deeper orange color (20). The relatively large area for β -cryp-

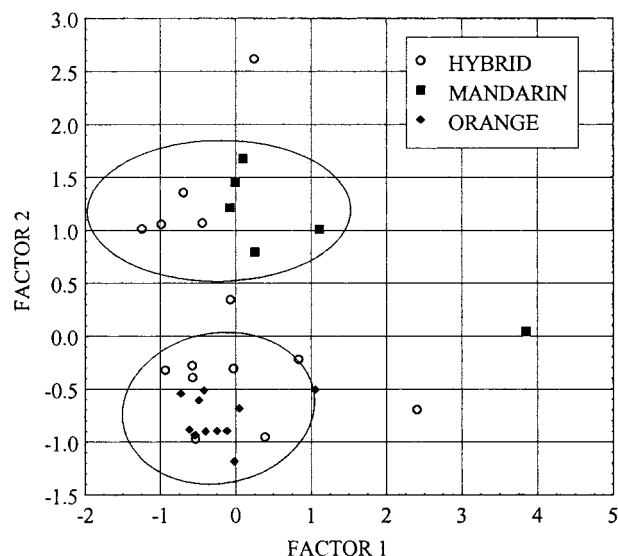


Figure 3. Results of applying PCA to the data. Suggested grouping ellipses are drawn.

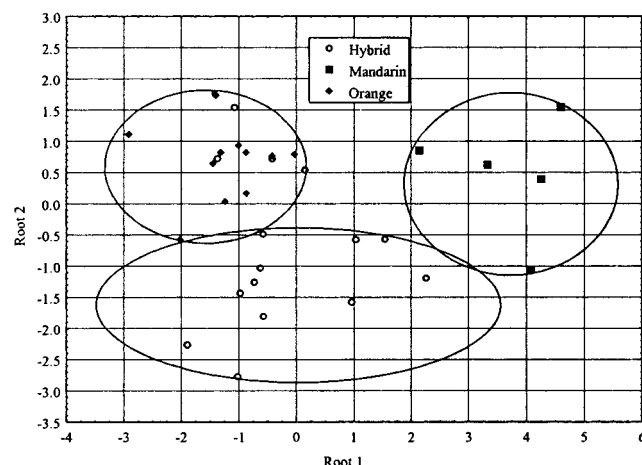
toxanthin in Rhode Red orange juice is still smaller than the smallest of the mandarins. This indicates that β -cryptoxanthin will very likely play an important role in the classification functions for mandarins and oranges.

However, β -cryptoxanthin cannot be the sole factor in classifying the juices by variety. To accomplish this a number of statistical methods were used. Principal components analysis (PCA) was used to examine the inherent structure of the data. The results from this analysis are shown in Figure 3 with the factor loadings shown in Table 3. It is apparent that for the most part the mandarins and oranges are separated very well with the exception of the Rhode Red orange juice. The hybrid varieties are scattered throughout the graph, indicating that some are mandarin-like, some are orange-like, and some are in a class by themselves. It can be seen that if groupings are drawn in a logical manner, many of the hybrids are grouped with either mandarins or oranges, along with an orange and a mandarin being grouped incorrectly. This implies that the PCA will not be successful in determining a model for differentiating the mandarins and oranges.

Due to the lack of adequate grouping using PCA, discriminate analysis was conducted. A graph of the data after canonical discriminate analysis (CDA) is shown in Figure 4 (loadings in Table 4) along with possible grouping ellipses. From this graph it is apparent that the orange and mandarin juices can be dif-

Table 3. PCA Eigenvector Loading Values

peak	factor 1	factor 2	factor 3
1	-0.026	0.577	0.567
2	0.297	0.039	0.814
3	0.748	-0.165	0.475
4	0.755	0.383	0.375
5	0.850	0.123	0.441
6	0.536	-0.123	0.590
7	0.863	0.236	0.143
8	0.115	0.967	0.009
9	0.880	-0.046	0.313
10	0.082	0.938	-0.040
11	0.946	-0.047	0.156
12	0.686	0.329	-0.255
variance	42.80	21.00	17.50
total	42.80	63.80	81.30

**Figure 4.** Results of applying CDA to the data. Suggested grouping ellipses are drawn.**Table 4. Factor Loadings Used in Creation of CDA Plot Shown in Figure 4 along with Means and Standard Deviations of the 12 Peaks**

	root 1	root 2	mean	SD
VAR1	-0.34	-0.53	0.0007	0.0000
VAR2	-0.48	-0.61	0.0026	0.0018
VAR3	0.23	0.61	0.0041	0.0033
VAR4	0.92	-0.78	0.0014	0.0010
VAR5	0.99	2.52	0.0141	0.0125
VAR6	0.01	-0.80	0.0126	0.0138
VAR7	0.57	0.70	0.0165	0.0190
VAR8	0.92	-0.83	0.0038	0.0039
VAR9	-2.99	-1.19	0.0028	0.0030
VAR10	-0.32	0.43	0.0717	0.0798
VAR11	1.34	-1.14	0.0012	0.0017
VAR12	-0.33	0.52	0.0065	0.0092
eigenvalue	3.00	0.57		
proportion of total	0.84	1.00		

ferentiated using the 12 chosen peaks. The hybrid varieties can be clearly distinguished from the mandarins and for the most part from the oranges. Of the 14 hybrid juices analyzed, 3 would be definitely classified as oranges and 2 are on the arbitrary orange border.

Another useful discriminate analysis method is to generate discriminate functions for the variables and then inspect the Mahalanobis distances to determine the classification. This has certain advantages over CDA, namely, the ability to analyze two groups. The Mahalanobis distances were used to discriminate between orange and mandarin juices with 100% accuracy. In fact, 100% accuracy could be achieved with only two variables, mutatoxanthin A and peak 8. One might expect, upon casual observation, β -cryptoxanthin to play

a major role in the discrimination of the two juices. In fact, β -cryptoxanthin can be used as a discriminating factor, which resulted in 88.2% classification accuracy. Several of the compounds are highly correlated: 0.848 for mutatoxanthin A (peak 4) and isolutein (peak 7), 0.929 for α -cryptoxanthin and α -carotene, and 0.927 for peak 8 and β -cryptoxanthin. It should be noted that all of these compounds have larger concentrations in mandarins versus oranges.

CONCLUSIONS

By using carotenoid profiles to develop models, it is possible to classify juice form as mandarin, orange, or hybrid fruit with high accuracy (>90%). There is the possibility of improving the classification by including more than just carotenoid profiles. For example, a model consisting of both carotenoids, such as the work presented here, and flavanone glycosides, as reported by Mouly and co-workers in 1994 (15), could provide better discrimination. The two techniques combined would provide a more complete and useful model.

ACKNOWLEDGMENT

R.L.R. acknowledges H.J.H.'s hospitality and cooperation during his time at the Gesellschaft für Lebensmittel-Forschung in Berlin, Germany. We are indebted to Dr. Gerry Spinwall of Hoffmann-La Roche for the kind gift of authentic carotenoid standards and to the YMC Corp. for the gift of chromatographic supplies.

LITERATURE CITED

- Gross, J. *Pigments in Fruits*; Harcourt Brace Javanovich: London, U.K., 1987.
- Wheaton, T. A.; Stewart, I. Optimum temperature and ethylene concentrations for postharvest development of carotenoid pigments in citrus. *J. Am. Soc. Hortic. Sci.* **1973**, *98*, 337–340.
- Curl, A. L.; Bailey, G. F. The carotenoids of Navel oranges. *J. Food Sci.* **1961**, *26*, 442–447.
- Yokoyama, H.; White, M. U. Citrus carotenoids. VI. Carotenoid pigments in the flavedo of Sinton citrangequat. *Phytochemistry* **1966**, *5*, 1159–1173.
- Bernath, P.; Swisher, H. E. Rapid quantitative method for determining carotenoids of the California sweet orange. *Food Technol. (Champaign)* **1969**, *23*, 815–818.
- Gross, J.; Gabai, M.; Lifshitz, A. Carotenoids in juice of Shamouti orange. *J. Food Sci.* **1971**, *36*, 466–473.
- Stewart, I.; Wheaton, T. A. Continuous flow separation of carotenoids by liquid chromatography. *J. Chromatogr.* **1971**, *55*, 325–336.
- Stewart, I.; Wheaton, T. A. Conversion of β -citaurin to reticulataxanthin and β -apo-8'-carotenal to citranaxanthin during the isolation of carotenoids from citrus. *Phytochemistry* **1973**, *12*, 2947–2951.
- Mouly, P. P.; Gaydou, E. M.; Lapiere, L.; Corsetti, J. Differentiation of several geographical origins in single-strength Valencia orange juice using quantitative comparison of carotenoid profiles. *J. Agric. Food Chem.* **1999**, *47*, 4038–4045.
- Sawamura, M.; Shichiri, K. I.; Ootani, Y.; Zheng, X. H. Volatile constituents of several varieties of pummelos and characteristics among citrus species. *Agric. Biol. Chem.* **1991**, *55*, 2571–2578.
- Rouseff, R. L.; Ting, S. V. Tangeretin content of Florida citrus peel as determined by HPLC. *Proc. Fla. State Hortic. Soc.* **1979**, *92*, 145–148.
- Gaydou, E. M.; Bianchini, J. P.; Randriamiharisoa, R. P. Orange and mandarin peel oils differentiation using polymethoxylated flavone composition. *J. Agric. Food Chem.* **1987**, *35*, 525–529.

- (13) Rouseff, R. Differentiating citrus juices using flavone glucoside concentration profiles. In *Adulteration of Fruit Juice Beverages*; Nagy, S., Attaway, J. A., Rhodes, M., Eds.; Dekker: New York, 1987; pp 49–65.
- (14) Rouseff, R. L. Liquid-chromatographic determination of naringin and neohesperidin as a detector of grapefruit juice in orange juice. *J. Assoc. Off. Anal. Chem.* **1988**, *71*, 798–802.
- (15) Mouly, P. P.; Arzouyan, C. R.; Gaydou, E. M.; Estienne, J. M. Differentiation of citrus juices by factorial discriminant analysis using liquid chromatography of flavanone glycosides. *J. Agric. Food Chem.* **1994**, *42*, 70–79.
- (16) Pino, J. A.; Torricella, R. G.; Orsi, F.; Figueras, L. Application of multivariate statistics for the quality classification of single-strength grapefruit juice. *J. Food Qual.* **1986**, *9*, 205–216.
- (17) Jella, P.; Rouseff, R. L.; Goodner, K. L.; Widmer, W. W. Determination of key flavor components in methylene chloride extracts from processed grapefruit juice. *J. Agric. Food Chem.* **1998**, *46*, 242–247.
- (18) Shaw, P. E.; Buslig, B. S.; Moshonas, M. G. Classification of orange and grapefruit juices by pattern recognition techniques. *Fluess. Obst* **1994**, *61*, 45–49.
- (19) Rouseff, R. L.; Raley, L.; Hofsommer, H. J. Application of diode array detection with a c-30 reversed phase column for the separation and identification of saponified orange juice carotenoids. *J. Agric. Food Chem.* **1996**, *44*, 2176–2181.
- (20) Stewart, I.; Bridges, G. D.; Peiringer, A. P.; Wheaton, T. A. 'Rhode Red Valencia,' an orange selection with improved juice color. *Fla. State Hortic. Soc.* **1975**, *88*, 17–19.
- (21) Farin, D.; Ikan, R.; Gross, J. The carotenoid pigments in the juice and flavedo of a mandarin hybrid (*Citrus reticulata*) cv. Michal during ripening. *Phytochemistry* **1983**, *22*, 403–408.
- (22) Van Heukelem, L.; Lewitus, A. J.; Kana, T. M. High-performance liquid chromatography of phytoplankton pigments using a polymeric reversed-phase C18 column. *J. Phycol.* **1992**, *28*, 867–872.
- (23) De Ritter, E.; Purcell, A. E. Carotenoid Analytical Methods. In *Carotenoids as Colorants and Vitamin A Precursors*; Bauernfeind, J. C., Ed.; Academic Press: London, U.K., 1981; pp 815–923.
- (24) Gross, J.; Gabai, M.; Lifshitz, A. A comparative study of the carotenoid pigments in juice of Shamouti, Valencia and Washington oranges, three varieties of *Citrus sinensis*. *Phytochemistry* **1972**, *11*, 303–308.

Received for review July 12, 2000. Revised manuscript received December 12, 2000. Accepted December 14, 2000. Mention of a trademark or proprietary product is for identification only and does not imply a guarantee or warranty of the product by the U.S. Department of Agriculture. The U.S. Department of Agriculture prohibits discrimination in all its programs and activities on the basis of race, color, national origin, gender, religion, age, disability, political beliefs, sexual orientation, and marital or family status.

JF000866O