JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Technological Classification of Basque Cider Apple Cultivars According to Their Polyphenolic Profiles by Pattern Recognition Analysis

ROSA M. ALONSO-SALCES,† CARLOS HERRERO,‡ ALEJANDRO BARRANCO,† LUIS A. BERRUETA,*,† BLANCA GALLO,† AND FRANCISCA VICENTE†

Departamento de Química Analítica, Facultad de Ciencias, Universidad del País Vasco/Euskal Herriko Unibertsitatea, P.O. Box 644, E-48080 Bilbao, Spain, and Departamento de Química Analítica, Nutrición y Bromatología, Facultad de Ciencias, Universidad de Santiago de Compostela, Augas Férreas s/n, Campus Universitario, 27002 Lugo, Spain

The polyphenolic compositions of 31 Basque cider apple cultivars were determined in pulp, peel, and juice by high-performance liquid chromatography−diode array detection analysis of crude extracts and after thiolysis. Data sets, consisting of individual polyphenol concentrations, total procyanidin content, and the average degree of polymerization of procyanidins, were evaluated by multivariate chemometric techniques, to develop decision rules for classifying apple cultivars technologically into bitter and nonbitter categories. A preliminary study of the data structure was performed by cluster analysis and principal component analysis in each apple material. Bitter apple varieties presented higher contents of flavan-3-ols and/or dihydrochalcones than nonbitter cultivars. Different classification systems for the two categories on the basis of the chemical data were obtained applying several supervised pattern recognition procedures, such as linear discriminant analysis, K-nearest neighbors, soft independent modeling of class analogy, partial least-squares, and multilayer feed forward artificial neural networks. Excellent performance in terms of recognition and prediction abilities for both categories (100% of hits) was achieved in every case (pulp, peel, or juice). Polyphenolic profiles of apple pulp, peel, or juice provide enough information to develop classification criteria for establishing the technological group of apple cultivars (bitter or nonbitter).

KEYWORDS: Polyphenol; apple; cider; bitterness; HPLC; thiolysis; pattern recognition

INTRODUCTION

Several works have been developed to study the chemical constituents and the technological qualities of different apple cultivars in order to select the most appropriate for the elaboration of ciders, juices, and other apple-derived products (*1*, *2*). The main technological properties that apple cultivars used for cider making should present are the following: a high juice yield; a medium-high level density and sugar content and a reduced dried extract; a balanced concentration of pectins, polyphenols, and organic acids; a low nitrogen content; aromas; and interesting sensory qualities. Moreover, it is required that apple fruit has a good resistance to manipulation during harvest and transportation and that maturation is late enough to process the fruits when temperatures are sufficiently low so that the fermentative process develops more slowly (*3*).

Apples present a wide diversity of polyphenols classified into several major classes. The flavan-3-ols include monomeric (catechins) and polymeric (procyanidins) forms, mainly constituted by (−)-epicatechin (EC) units. Among the hydroxycinnamic acids, caffeoylquinic acid and *p*-coumaroylquinic acid show the highest contents. The major species of the dihydrochalcones are phloretin glucoside and xyloglucoside, which are generally considered specific to apples. Last, flavonols and anthocyanins are essentially present in apple peels (*4*). The interest in polyphenols in cider apple cultivars is due to the fact that they are responsible for the color and the balance of bitterness to astringency, which defines the "overall mouth feel" of the ciders (*5*). Furthermore, they are implicated in alcoholic and malolactic fermentations as metabolites, providing cider aroma, and as inhibitors of microbiological growth, controlling fermentation rates and cider spoilage (*6*). Polyphenols are also involved in the colloidal stability of cider (*7*).

Technological classification of cider apple varieties is commonly based on the total polyphenol content and the total acidity of their juices. Following these criteria, apple cultivars are classified in six technological groups: sweet (<3.55 g sulfuric acid/L, <1.45 g tannic acid/L), bittersweet (<3.55 g sulfuric acid/L, >1.45 g tannic acid/L), semiacid (3.55−4.80 g sulfuric acid/L, <1.45 g tannic acid/L), semiacid-bitter (3.55−4.80 g

* To whom correspondence should be addressed. E-mail: qapbesil@lg.ehu.es.
† Universidad del País Vasco/Euskal Herriko Unibertsitatea.
‡ Universidad de Santiago de Compostela.

10.1021/jf048995i CCC: $27.50
Published on Web 12/01/2004

sulfuric acid/L, >1.45 g tannic acid/L), acid (>4.80 g sulfuric acid/L, <1.45 g tannic acid/L), and acid-bitter (>4.80 g sulfuric acid/L, >1.45 g tannic acid/L) (*3*). However, the information obtained by the analysis of these global parameters is limited because any distinction among the different classes of polyphenols and their diverse properties is made. This kind of information is especially interesting taking into account that some polyphenols or classes are those that give certain characteristics to the final product. Thus, hydroxycinnamic acids are precursors of volatile phenols formed during fermentation that contributes to cider aroma (*8*). Caffeoylquinic acid and catechins generate colored products by enzymatic oxidation and coupled oxidation reactions with other polyphenols (*9*). Procyanidins are responsible for cider bitterness and astringency (*10*).

In addition, these global estimations performed on juice do not provide complete information on the polyphenolic potential of fruit, since an important part of the native compounds is oxidized and adsorbed on apple cell walls when juice is made. Therefore, a precise knowledge of the composition of cider apple cultivars may contribute to a better understanding of their implication in the quality and diversity of apple-derived products, such as cider and apple juice. In this sense, several characterization studies of different dessert apple varieties (*11*, *12*) and cider apple cultivars from Spain (*13*, *14*), France (*4*), and the United Kingdom (*15*) have been carried out on the basis of their polyphenolic profiles.

Cider should be made with a mixture of different cider apple cultivars, that is, cultivars from the different technological groups, to obtain an apple juice with a balanced composition in the substance of technological interest. These components allow an adequate fermentation process and give the juice certain characteristics related to flavor, color, product stability, microbiological control, etc., to attain a quality cider with particular organoleptic properties. The aim of this work is to achieve classification rules for predicting the technological group (bitter or nonbitter) to which Basque cider apple cultivars belong, according to their polyphenolic profiles.

MATERIALS AND METHODS

Reagents and Standards. Methanol (Romil Chemical Ltd, Heidelberg, Germany) was of high-performance liquid chromatography (HPLC) grade. Water was purified on a Milli-Q system from Millipore (Bedford, MA). Glacial acetic acid, formic acid, toluene-α-thiol, Folin−Ciocalteu reagent, fuming hydrochloric acid 37%, sodium hydroxide, and potassium hydrogen phthalate (GR volumetric standard) were provided by Merck (Darmstadt, Germany), and ascorbic acid by Panreac (Barcelona, Spain) was of analytical quality. All solvents used were previously filtered through 0.45 μm nylon membranes (Lida, Kenosha, WI).

Polyphenol standards were supplied as follows: (+)-catechin (CAT), EC, rutin (RUT), phloridzin (PLG), 5-caffeoylquinic acid (CQA), *p*-coumaric acid, and tannic acid by Sigma-Aldrich Chemie (Steinheim, Germany); hyperoside (HYP), isoquercitrin (IQC), avicularin (AVI), quercitrin (QCI), and ideain (IDE) chloride by Extrasynthèse (Genay, France). EC 4*R*-benzylthioether and 4-*p*-coumaroylquinic acid (PCQ) were kindly provided by Dr. S. Guyot (INRA, France), and phloretin-2′-xyloglucoside and procyanidin B2 were provided by Dr. F. A. Tomás-Barberán (CEBAS-CSIC, Spain) and Dr. C. Santos-Buelga (Universidad de Salamanca, Spain), respectively. Stock standard solutions of CAT, EC 4*R*-benzylthioether, EC, RUT, PLG, CQA, and *p*-coumaric acid at a concentration of 1 mg mL^{-1} and HYP, IQC, QCI, and IDE at 0.6 mg mL^{-1} were prepared in methanol and stored at 4 °C in darkness. The other standards were prepared in approximate concentrations and used for chromatographic peak identification.

Plant Materials. Pulps and peels from 31 different apple cultivars used in the Basque Country for cider making were analyzed as follows: Azpuru Garratza (AG), Bost Kantoi (BK), Errezila (ER), Gazigorri (GG), Goikoetxea (GK), Geza Miña (GM), Gazilokia (GZ), Ibarra (IB), Larrabetzu (LR), Manttoni 111 (MN111), Manttoni EM7 (MNEM7), Moko (MK), Mendexa 1 (MX1), Mendexa 10 (MX10), Mendexa 11 (MX11), Mendexa 2 (MX3), Mendexa 3 (MX2), Mendexa 4 (MX4), Mozoloa (MZ), Piko (PK), Palazio (PL), Patzuloa (PT), Txistu (TT), Txalaka (TX), Ugarte (UG), Urdai Goika Santutxu (UGS), Udare Marroi (UM), Urtebi Haundia (UH), Urdin (UR), Urdin Zalla (URZ), and Urtebi Txiki (UT). Apples were harvested in the Experimental Orchards of the Diputación Foral de Gipuzkoa in Hondarribia (Guipúzcoa, Spain) and the Diputación Foral de Bizkaia in Zalla (Vizcaya, Spain) during the 2000 and 2001 seasons.

Apple Powder Preparation. Fruits were harvested at maturity, which was tested by the lugol index (*16*). For each variety and season, two or three batches of 10 apple fruits were mechanically peeled and cored and sprayed with an aqueous solution of 3% formic acid in order to avoid polyphenol oxidation. Peels and pulps were immediately frozen in liquid nitrogen, and then, they were freeze-dried. An aliquot for each variety was used to determine the fresh/dry matter ratio. The dried tissues were crushed in closed vials to avoid hydration, obtaining a homogeneous powder that was stored at room temperature in a desiccator until analysis. Aliquots of 0.5 g of freeze-dried apple peels or pulps were used for each analysis.

Apple Juice Preparation. For preparing juice samples, two or three batches of fruits (1 kg) were constituted for each cultivar. Each batch was milled and pressed to obtain crude juice, using procedures similar to those used by Basque cider makers (a grinder and a traditional press) but in small scale. A solution of diluted sodium fluoride (50 mL, 1 g/L in water) was added to the apples before pressing to avoid oxidation to a certain extent. This added volume was subtracted for yield calculations, and a correcting factor was applied for calculating polyphenol concentrations. Then, crude apple juices were centrifuged (10000 rpm, 15 min) at 4 °C to obtain clear apple juices. Aliquots of centrifuged apple juices were sampled for the determination of polyphenolic profiles by HPLC (2 × 1 mL), total polyphenol content by the Folin−Ciocalteu method (0.5 mL), and total acidity (40 mL). Aliquots for HPLC analyses were freeze-dried and stored in a desiccator until analysis. Other aliquots were frozen and kept at −20 °C and were defrosted just before analysis.

Analytical Procedures. *Thiolysis and Direct Solvent Extraction and Reversed Phase HPLC Analysis of Freeze-Dried Samples.* Freeze-dried samples were submitted to thiolysis as described by Guyot et al. (*17*) and to direct solvent extraction with 30 mL of methanol−water−acetic acid (30:69:1, v/v/v) with ascorbic acid (2 g/L) in an ultrasonic bath during 10 min (*18*). Then, both thiolysis reaction mixtures and crude solvent extracts were filtered through a 0.45 μm PTFE filter (Waters, Milford, CA) prior to injection into the HPLC system.

Chromatographic analysis was performed on a Hewlett-Packard Series 1100 system, equipped with a vacuum degasser, a quaternary pump, a thermostated autosampler, a thermostated column compartment, and a diode array detector (DAD), connected to HP ChemStation software. A reversed phase Nova-Pak C18 (300 mm × 3.9 mm i.d., 4 μm) column and a Nova-Pak C18 (10 mm × 3.9 mm i.d., 4 μm) guard column (Waters, Barcelona, Spain) were used. Solvents that constituted the mobile phase were acetic acid−water, 10:90, v/v (A), and methanol (B). The elution conditions applied were as follows: 0−10 min, 0% B isocratic; 10−40 min, linear gradient 0−15% B; 40−60 min, 15% B isocratic; and finally, washing and reconditioning of the column. The flow rate was 0.8 mL min^{-1}, and the injection volume was 50 μL of the crude extracts or 10 μL of the thiolysis media. The system was operated at 25 °C. Flavan-3-ols and dihydrochalcones were monitored and quantified at 280 nm, hydroxycinnamic acids at 320 nm, flavonols at 370 nm, and anthocyanins at 530 nm. Polyphenol identification for which standards were available was carried out by comparison of their retention times and their UV−visible spectra with those of the standards. Some other chromatographic peaks were assigned to a particular polyphenol class according to their UV−visible spectra and bibliographic sources. In this sense, those unknown chromatographic peaks that exhibit flavan-3-ol spectra were appointed as CAT-*n*, and those with a spectrum of CQA were appointed as unknown hydroxycinnamic acids with caffeic acid UV spectra (CAA-*n*), of *p*-coumaric as CMA-

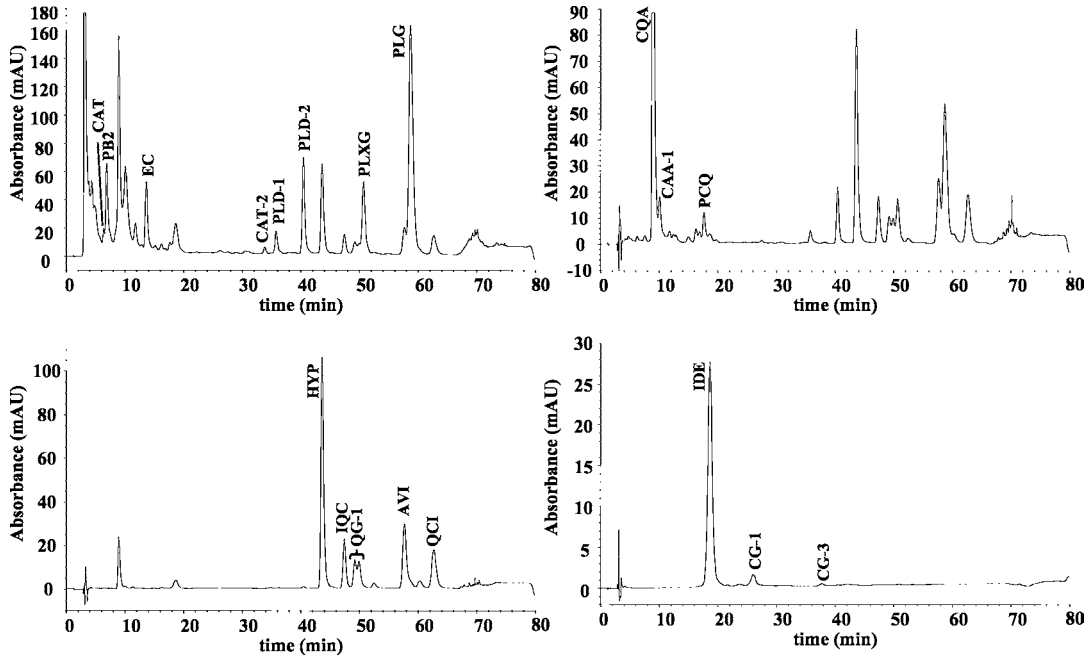

Figure 1. Chromatograms of apple peel polyphenols, obtained by HPLC-DAD at the different wavelengths used for quantitation: (**a**) 280, (**b**) 320, (**c**) 370, and (**d**) 530 nm. Apple cultivar, GK.

n, of dihydrochalcone as PLD-*n*, of flavonol as QG-*n*, and of anthocyanin as CG-*n* (where "*n*" is a number). Quantification was performed by reporting the measured integration areas in the calibration equation of the corresponding standards. Thus, procyanidin B2 and the unknown flavan-3-ols (CAT-2) were quantified as CAT; phloretin-2′-O-xyloglucoside (PLXG) and the unknown dihydrochalcones were quantified as PLG; AVI and the unknown flavonols were quantified as RUT; CAA-*n* species were quantified as CQA; PCQ and CMA-*n* species were quantified as *p*-coumaric acid; and the unknown anthocyanins (CGs) were quantified as IDE.

Total Polyphenol Content by Folin−Ciocalteu Method. Estimation of the global polyphenol content in apple juices was performed according to the Folin−Ciocalteu method adapted from Singleton and Rossi (*19*). Centrifuged juice aliquots (0.5 mL) were diluted 20-fold in methanol−acetic acid 2.5% (10:90). Folin−Ciocalteu reagent (0.25 mL) was added to 0.5 mL of the diluted cider solution. The mixture was homogenized with a vortex and after the reaction took place for 3 min, 1 mL of Na_2CO_3 (200 g/L) and 3.25 or 8.25 mL of ultrapure water were added, depending on the cultivar polyphenol content, and homogenized. Then, the mixture was incubated for 10 min at 70 °C. Once it had cooled at room temperature, it was homogenized and its absorbance was measured at 700 nm with a Shimadzu UV-260 spectrophotometer (Kyoto, Japan) against a blank [0.5 mL of methanol−acetic acid 2.5% (10:90) plus reagents] in the reference cell. Quantification was achieved by reporting the absorbances in the calibration curve of tannic acid used as a standard polyphenol.

Total Acidity and pH of Apple Juices. The apple juice total acidity was determined by a potentiometric titration. An aliquot of apple juice (40 mL) was placed in a glass cell, as well as 40 mL of water that had been previously boiled and cooled at room temperature. An aqueous solution of NaOH (0.1 M) was used as tritator, once it had been standardized with potassium hydrogen phthalate. The automated system used to perform the potentiometric titration was developed by Cazallas et al. (*20*), using a Ag−AgCl(s) reference electrode and a glass electrode. Titrator additions were carried out with an automatic buret Metrohm Dosimat 725. The whole system was controlled by the software POSPETR (*20*). Analyses were performed at 25 °C. The titration equivalence point was calculated by considering the titrator added volume and the potential measurements in each addition using the software POTCAL (*21*). Total acidity results were expressed in grams of sulfuric acid per liter of juice. Apple juice pH values were measured with a Mettler Toledo MP-125 pH meter (Greifensee, Switzerland).

Table 1. Concentrations (mg/kg of Apple) of Flavan-3-ols, Hydroxycinnamic Acids, Dihydrochalcones, and Flavonols and the DPn in Apple Pulps (2000 and 2001 Seasons)[a]

season	2000				2001			
n	22				30			
polyphenol	mean	SD	min	max	mean	SD	min	max
Flanvan-3-ols								
CAT	38	29	4	115	44	74	0.7	407
EC	142	103	24	398	156	146	39	770
PB2	134	119	28	529	143	120	44	514
CAT-2	12	10	3	46	13	10	5	43
PC	1791	2050	675	10388	1575	816	672	4448
DPn	5	1	4	8	4.6	0.6	3.5	6.0
Hydroxycinnamic Acids								
CQA	426	486	67	2420	335	178	61	724
PCQ	24	24	0.5	79	26	28	1	120
CAA-1	23	16	5	63	24	16	ND	64
CMA-2	0.8	0.7	0.0	2.8	0	1	ND	7
Dihydrochalcones								
PLXG	23	14	7	54	28	19	4	67
PLG	23	33	5	159	17	13	5	60
PLD-1	5	4	ND	15	3	2	0.6	9
PLD-2	5	6	ND	24	3	2	ND	11
Flavonols								
HYP	0.1	0.2	ND	0.9	0.02	0.09	ND	0.47
IQC	0.5	0.8	ND	3.2	0.4	0.4	ND	1.8
QCI	2	2	ND	5	2	1	0.3	6
QG-1	1	1	ND	6	0.6	0.4	ND	1.6

[a] *n*, number of cultivars studied.

Data Analysis and Chemometric Procedures. The apple peel data set consisted of a 70 × 27 matrix, the apple pulp data set consisted of a 97 × 18 matrix, and the apple juice data set consisted of a 97 × 19 matrix. Rows represented apple samples, and columns represented the concentration of individual polyphenols determined by HPLC-DAD, the total concentration of procyanidins, and the average degree of polymerization of procyanidins (DPn). Each sample was represented in the multidimensional space by a data vector, which is an assembly of the 27 features in peel, the 18 features in pulp, and the 19 features in juice. In each apple material, separately, data vectors belonging to the same category (bitter or nonbitter) were analyzed using chemometric

Table 2. Concentrations (mg/kg of Apple) of Flavan-3-ols, Hydroxycinnamic Acids, Dihydrochalcones, Flavonols, and Anthocyanins and the DPn in Apple Peels (2000 and 2001 Seasons)

season	2000				2001			
n	22				30			
polyphenol	mean	SD	min	max	mean	SD	min	max
Flanvan-3-ols								
CAT	7	6	1	29	7	9	0.3	41
EC	52	44	9	196	61	60	11	302
PB2	56	52	12	252	46	31	8	122
CAT-2	6	5	2	25	5	3	2	14
PC	877	792	283	4058	777	314	360	1676
DPn	6	1	4	9	5.6	0.7	4.6	7.1
Hydroxycinnamic Acids								
CQA	56	123	2	593	37	33	5	139
PCQ	3	4	ND	18	3	4	ND	17
CAA-1	4	3	0.9	17	6	4	0.8	22
CAA-2	1.0	0.9	0.2	3.4	1	1	ND	4
CMA-2	0.6	0.6	ND	2.1	0.6	0.8	ND	3.1
Dihydrochalcones								
PLXG	16	11	4	56	17	12	2	60
PLG	54	80	12	378	41	35	7	123
PLD-1	5	3	1	13	4	3	1	12
PLD-2	12	13	2	56	8	6	0.6	24
Flavonols								
HYP	23	15	1	60	29	16	4	68
IQC	8	4	1	21	9	6	2	25
AVI	18	13	6	67	20	10	7	52
QCI	9	5	1	25	11	7	3	29
QG-1	12	7	3	33	13	6	ND	28
QG-2	1	1	0.2	6	2	1	0.3	6
QG-3	0.4	1	ND	4	0.5	1	ND	4
Anthocyanins								
IDE	2	5	ND	25	2	3	ND	10
CG-1	0.08	0.2	ND	0.8	0.06	0.1	ND	0.5
CG-2	ND				0.005	0.02	ND	0.08
CG-3	0.03	0.1	ND	0.6	0.02	0.05	ND	0.22
CG-4	0.03	0.1	ND	0.6	0.01	0.04	ND	0.21

Table 3. Concentrations (mg/kg of Apple) of Flavan-3-ols, Hydroxycinnamic Acids, Dihydrochalcones, and Flavonols and the DPn in Apple Juices (2000 and 2001 Seasons)

season	2000				2001			
n	17				27			
polyphenol	mean	SD	min	max	mean	SD	min	max
Flanvan-3-ols								
CAT	37	23	4	91	44	79	1	407
EC	91	33	49	154	169	174	19	822
PB2	82	25	46	123	151	131	38	550
CAT-2	6	2	3	9	8	12	ND	48
PC	651	123	413	896	975	840	281	3511
DPn	3.5	0.3	2.8	4.2	3	0.5	3	5
Hydroxycinnamic Acids								
CQA	371	216	101	820	497	264	91	1099
PCQ	46	43	7	147	55	74	3	282
CAA-1	17	7	7	32	6	16	ND	61
CAA-2	3	2	0.6	7	3	3	ND	8
CMA-2	3	4	ND	13	0.6	1	ND	5
Dihydrochalcones								
PLXG	34	18	11	85	46	32	10	137
PLG	28	20	11	88	25	22	11	95
PLD-1	1.4	0.9	ND	3.1	2	2	ND	8
PLD-2	1.1	0.7	ND	3.1	0.9	1	ND	3
Flavonols								
HYP	0.6	0.1	0.3	0.8	1.3	0.7	0.4	4.0
IQC	0.6	0.2	0.3	1.1	0.9	0.8	0.2	4.1
QCI	1.3	0.7	0.5	3.0	2	2	0.6	6
QG-1	1.2	0.5	0.5	2.3	0.4	0.6	ND	2.4

procedures that have been described in the literature (*22*, *23*), such as cluster analysis (CA), principal component analysis (PCA), linear discriminant analysis (LDA), K-nearest neighbors (KNN), soft independent modeling of class analogy (SIMCA), partial least-squares (PLS), and multilayer feed forward artificial neural networks (MLF-ANN). Statistical and chemometric data analyses were performed by means of the statistical software packages Statgraphics (*24*), Parvus (*25*), SPSS (*26*), and The Unscrambler (*27*).

CA is a pattern recognition technique that is used to reveal the structure residing in a data set and disclose the natural groupings existing between samples characterized by the values of a set of measured variables. It is commonly applied before other multivariate techniques because of its unsupervised character. CA was performed on the autoscaled data. Sample similarities were calculated on the basis of the squared Euclidean distance, and the Ward hierarchical agglomerative method was used to establish clusters (*28*).

PCA, performed on the autoscaled data, allowed us to reduce the number of variables retaining the maximum amount of variability present in the data in order to provide a partial visualization of data structure in a reduced dimension.

The supervised pattern recognition techniques LDA, KNN, SIMCA, PLS, and a MLF neural network were used in order to attain classification rules (bitter/nonbitter) for predicting the bitterness of apple cultivars according to their polyphenolic profiles. These techniques were applied to the autoscaled data matrices of each apple material (pulp, peel, and juice). A cross-validation of the classification rules proposed by these techniques was performed by dividing the complete data set into a training set and an evaluation set. Samples were assigned randomly to a training set consisting of 75% of them, and the test set was composed of the remaining 25% of samples. Such a division allowed for a sufficient number of samples in the training set and a representative number of members among the test set. The same process was repeated four times with different constitutions of both sets to ensure that all of the samples had the possibility to be included in the evaluation set at least once. The different pattern recognition techniques were applied to the four training test sets obtained. The reliability of the classification models achieved was studied in terms of recognition ability (percentage of the members of the training set correctly classified) and prediction ability (percentage of the members of the test set correctly classified by using the rules developed in the training step).

SIMCA version applied was the normal range model. The stop criterion used by SIMCA for the calculation of the principal components was the minimum percentage of retained variance (95%, $\alpha = 0.05$). The model obtained by SIMCA for each category was also evaluated in terms of sensitivity and specificity. The sensitivity of the model is the percentage of objects belonging to the category, which are correctly identified by the mathematical model, and the specificity, the percentage of objects foreign to the category, which are classified as foreign (*29*).

When using KNN, the inverse square of the Euclidean distance was used as the criterion for calculating the distance between samples. The number of neighbors (K) was selected by studying the success in classification of this technique, when it was applied to a training set containing all of the samples and using different K values.

PLS analysis was performed using the polyphenolic profiles of each apple material as predictor variables and a binary response (0 = nonbitter category, and 1 = bitter category) as criterion variables. The total prediction error [prediction root mean square error (PRMSE)] and the residual error for every model sample were evaluated.

WinNN32 MLF-ANN (*30*) was applied on an input pattern consisting of the autoscaled data matrices of each apple material. The target output was assigned as 0 or (0, 1) for nonbitter cultivars and 1 or (1, 0) for bitter ones. A sigmoidal function $f(x) = 1/(1 + [\exp(-x)])$ was used as the transfer function. The neural network was trained by means of an algorithm that combined the use of an adaptative learning rate parameter (η) and a momentum (μ), which have been described previously (*22*). The initial values of the weights associated with the connections between neurons were selected randomly in the range -3 to 3. The maximum number of epochs was 2000, the initial values of η and μ were 0.2 and 0.5, respectively, and the target error was 0.1.

Table 4. Mean and SD of the Total Polyphenol Content (Folin–Ciocalteu Method) (g Tannic Acid/L) ($n = 3$) of Basque Cider Apple Juices in the 2000 and 2001 Seasons and pH and Total Acidity (g H_2SO_4/L) in Both Seasons Altogether

	2000 and 2001 seasons						total polyphenols			
	pH		total acidity				2000 season		2001 season	
cultivars	mean	SD	mean	SD	technological group[a]		mean	SD	mean	SD
GM	4.44	0.05	1.00	0.07	sweet	bitter			1.70	0.08
MX10	4.45	0.01	1.31	0.03	sweet	bitter			4.3	0.1
MX3	3.97	0.05	2.5	0.1	sweet	bitter			2.50	0.06
MZ	4.7	0.2	0.88	0.06	sweet	bitter	1.61	0.05	1.43	0.03
PK	4.4	0.1	1.2	0.3	sweet	bitter	1.4	0.1	1.55	0.08
PL	4.50	0.09	1.32	0.01	sweet	nonbitter	1.24	0.08	1.29	0.06
PT	4.3	0.3	1.0	0.1	sweet	bitter	1.64	0.05	1.29	0.06
UG	4.4	0.2	1.2	0.5	sweet	bitter	1.21	0.04	1.50	0.07
UGS	4.34	0.01	0.64	0.04	sweet	bitter			3.7	0.1
AG	3.62	0.06	3.97	0.02	semiacid	nonbitter	0.99	0.06	0.75	0.05
BK	3.54	0.04	4.3	1.2	semiacid	nonbitter	0.80	0.09	0.52	0.03
GZ	3.4	0.1	4.1	0.6	semiacid	nonbitter	1.09	0.04	0.80	0.07
IB	3.91	0.09	4.0	0.6	semiacid	nonbitter	0.82	0.02	1.0	0.1
MN111	3.7	0.1	3.6	0.5	semiacid	nonbitter	1.00	0.04		
MNEM7	3.8	0.2	4.0	0.7	semiacid	nonbitter	1.30	0.03	1.01	0.07
TT	3.62	0.02	4.2	0.3	semiacid	nonbitter	1.4	0.1		
UR	3.6	0.1	3.7	0.7	semiacid	nonbitter	1.53	0.05	0.83	0.03
URZ	3.62	0.03	4.4	0.5	semiacid	nonbitter	1.5	0.2	1.1	0.1
UT	3.6	0.1	4.22	0.05	semiacid	nonbitter			0.70	0.03
ER	3.54	0.01	5.8	0.2	acid	nonbitter			0.76	0.04
GG	3.34	0.05	4.9	0.3	acid	nonbitter			2.14	0.08
GK	3.54	0.09	4.8	0.8	acid	nonbitter	1.25	0.09	0.96	0.03
MK	3.22	0.07	9.0	0.5	acid	bitter			3.1	0.2
MX11	3.42	0.04	7.3	0.5	acid	nonbitter			1.23	0.06
MX2	3.23	0.04	8.3	0.1	acid	bitter			3.5	0.2
MX4	3.11	0.02	9.3	0.7	acid	nonbitter			1.25	0.09
TX	3.4	0.1	5.3	1.7	acid	nonbitter	0.9	0.1	0.80	0.05
UH	3.37	0.02	5.2	1.6	acid	nonbitter	0.95	0.06	0.79	0.03
UM	3.18	0.03	8.7	0.6	acid	nonbitter			1.0	0.1
LR	3.80[b]		1.98[b]		sweet	bitter	13.60[b]			
MX1	3.07[b]		7.34[b]		acid	bitter	3.63[b]			

[a] Technological classification of apple cultivars according to their total acidity and their polyphenolic profiles. Bitterness predictions are reported by Alonso-Salces et al. (*31*). [b] Personal communication of Dr. G. del Campo, Departamento de Química Aplicada, Universidad del País Vasco, San Sebastián, Spain.

RESULTS AND DISCUSSION

Polyphenolic profiles of cider apple cultivars in pulps, peels, and juices for the 2000 and 2001 harvests were characterized by HPLC-DAD. **Figure 1** shows an example of the chromatograms used for the quantitation of polyphenols. Analytical data are summarized in **Tables 1−3**. The detailed polyphenolic profiles, listing the individual polyphenol concentrations for each apple cultivar, were reported by Alonso-Salces et al. (*31*). In addition, the polyphenolic profile of each cultivar was related to its sensory properties (bitterness, astringency), its susceptibility to oxidation, and its possible influence on the characteristics and quality of the final product (cider, juice) when apples are processed (*31*).

For apple juices (17 cultivars of the 2000 season and 27 cultivars of the 2001 season), their total polyphenol contents, total acidity, and pH were also determined (**Table 4**).

Preliminary Statistic Data Treatment. In a first approach, an analysis of variance was performed on each apple material's (pulp, peel, and juice) data, considering only those varieties harvested the two seasons, to verify if there were significant differences in the individual polyphenol concentrations, total procyanidin (PC) contents, and DPn between both seasons. Most variables were not significantly different, except for some features that were present in very low concentrations (<2% of the total polyphenol contents). Moreover, box and whisker plots of these features confirmed that they present insufficient discriminatory abilities since they showed an overlap between the variable ranges in the two seasons. The differences observed in some features are likely due to the influence on fruit composition of certain factors, such as the weather, the nutrients status of the soil, and other environmental factors (*7*). Therefore, they were considered as part of the possible variability that apple compositions could present among seasons. Following, CA and PCA were carried out on the data of each apple material, but no natural groupings of the samples due to the harvest season were detected in pulps and peels. However, in juices, two partially overlapped groups were observed. Apple pulp and peel compositions were not significantly different between seasons, so the differences observed in apple juices were because of the slightly different method used each season to make the juices. Hence, juice data included the variability introduced by the juice elaboration procedure.

After this preliminary study, the complete data matrices of peels, pulps, and juices with all of the cider apple cultivars studied were considered, and their technological characterization was performed by classifying them as bitter or nonbitter on the basis of their polyphenolic profiles. The traditional classification of apple cultivars in technological groups, based on total acidity and total polyphenol content (Folin−Ciocalteu method) of the monovarietal apple juices, was used to establish the bitterness of each cultivar (**Table 4**). The data of those cultivars that were not clearly classified in one category depending on the harvest, or their apple juices were not available, were not considered to develop apple classification rules in the two established categories, bitter or nonbitter. Besides, those varieties, which in the preliminary CA were inside groups of the other category,

Table 5. Apple Cultivars Used for Developing Classification Rules with Apple Pulp, Peel, and Juice Data by Pattern Recognition Techniques

cultivar	class	apple material
AG	nonbitter	pulp, peel, juice
BK	nonbitter	pulp, peel, juice
ER	nonbitter	pulp, peel, juice
GK	nonbitter	pulp, juice
GZ	nonbitter	pulp, peel, juice
IB	nonbitter	pulp, peel, juice
MN111	nonbitter	pulp, peel, juice
MNEM7	nonbitter	pulp, peel, juice
MX11	nonbitter	pulp, juice
MX4	nonbitter	pulp, juice
PL	nonbitter	pulp, juice
TX	nonbitter	pulp, peel, juice
UH	nonbitter	pulp, peel, juice
UM	nonbitter	peel, juice
UR	nonbitter	pulp, peel, juice
URZ	nonbitter	pulp, peel, juice
UT	nonbitter	pulp, peel, juice
GG	bitter	juice
GM	bitter	pulp, peel
MK	bitter	pulp, peel, juice
MX10	bitter	pulp, peel, juice
MX2	bitter	pulp, juice
MX3	bitter	pulp, peel, juice
MZ	bitter	pulp, peel
PK	bitter	pulp, peel, juice
PT	bitter	pulp, peel
UG	bitter	pulp, peel, juice
UGS	bitter	pulp, juice

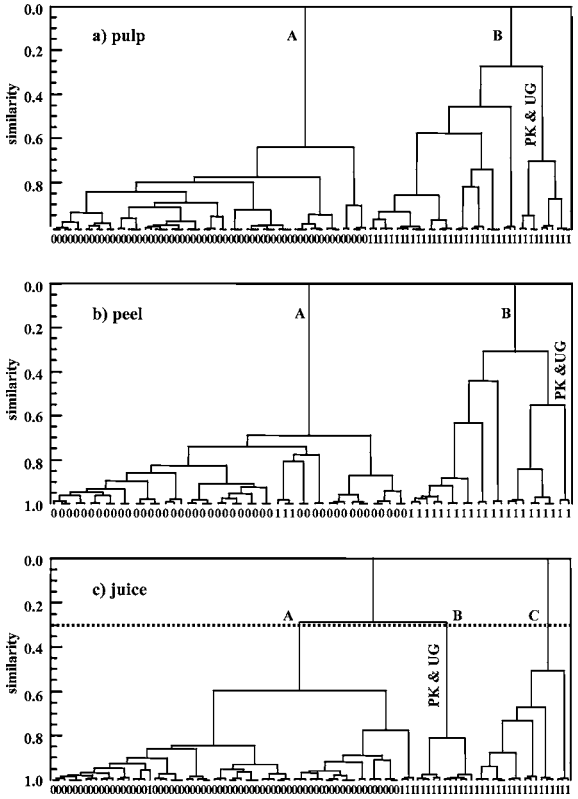

Figure 2. Dendrograms of CA for apple data: (**a**) pulp, (**b**) peel, and (**c**) juice. Sample codes: 0, nonbitter; 1, bitter.

were also not considered for this purpose. In **Table 5**, varieties used for developing the decision rules are shown. Once the classification rules were established, bitterness predictions of apple cultivars for which this information was not known or was confusing were performed.

Univariate Data Analysis. Despite the differences observed between bitter and nonbitter varieties when the individual polyphenol concentrations, PC content, and DPn in pulps, peels, and juices were considered, the box and whisker plots of these features in each apple material showed an overlap between the two classes, indicating insufficient discriminatory ability. Thus, none of the variables measured was able to discriminate between the established categories by itself. Hence, a multivariate approach was studied.

Multivariate Data Analysis. *CA.* CA results in each apple material are presented in the dendrograms of **Figure 2**. In pulps and peels, two clusters due to nonbitter (A) and bitter (B) samples, respectively, were observed. In peels, samples of the bitter variety UG (2000) were included in the nonbitter cluster. In juices, at a similarity level of 0.30, three clusters were identified as follows: cluster A, made up of nonbitter varieties and four bitter samples (MX3 and UG); cluster B, consisting of the bitter varieties PK and UG; and cluster C, containing the rest of the bitter cultivars. It is interesting to note the behavior of PK and UG cultivars that, being bitter, constituted a separate cluster at a similarity level of 0.30 in the three apple materials. These observations suggest that PK and UG varieties would present similar polyphenolic profiles between themselves but different from the other bitter varieties. This fact was not possible to be observed from the traditional apple classification in technological groups by measuring total polyphenol contents. In juices, PK and UG polyphenolic compositions are closer to the nonbitter class than to the bitter class, which can be explained taking into consideration that these varieties present intermediate polyphenol concentrations and are susceptible to suffering oxidation, because of their relatively high pH (4.4), close to the optimum pH for polyphenoloxidase (PPO) activity (4.5−5) (*32*). Therefore, polyphenol concentrations in their juices are relatively lower than those in the rest of the bitter varieties and more similar to nonbitter cultivars that are not so sensitive to oxidation.

PCA. In the tridimensional (pulp and peel) and bidimensional (juice) plots of the sample scores in the space defined by the three and two first principal components (PCs), respectively, a natural separation of bitter and nonbitter apple varieties is observed (**Figure 3**). Moreover, these plots revealed that the nonbitter class is much more homogeneous than the bitter class, because the latter was constituted by apple cultivars that presented very diverse polyphenolic profiles. PK and UG varieties conformed a subgroup inside the bitter class, as it occurred in CA. MX10 and MX2 cultivars contribute to the variability of the bitter class, showing up far from the rest of the bitter varieties. In juice, it can be observed that the MX3 cultivar is closer to PK and UG samples than to the other bitter ones. This bitter group made up of PK, UG, and MX3 presents polyphenolic compositions more similar to the nonbitter class than to the bitter class, as it was also disclosed by CA.

In pulps, the three PCs accounted for 68% of the total system variability. From the loadings of the variables (**Table 6**), catechins (EC and CAT) and the major dihydrochalcones (PLG and PLXG) were the ones that contributed more to the PC1. Features dominant in the second principal component (PC2) were procyanidins (PB2, CAT-2, and PC). In peels, PCA allowed us to reduce the number of variables from 27 to three, keeping 65% of the total information of the system. Flavan-3-ols (EC, CAT-2, CAT, and PB2) and major dihydrochalcones, PLXG and PLG, were the most influential features in PC1 (**Table 6**). In juices, the two PCs accounted for 61% of the total system variability. The major contribution to PC1 is due to flavan-3-ols (PC, PB2, CAT-2, and EC) and QCI, and the

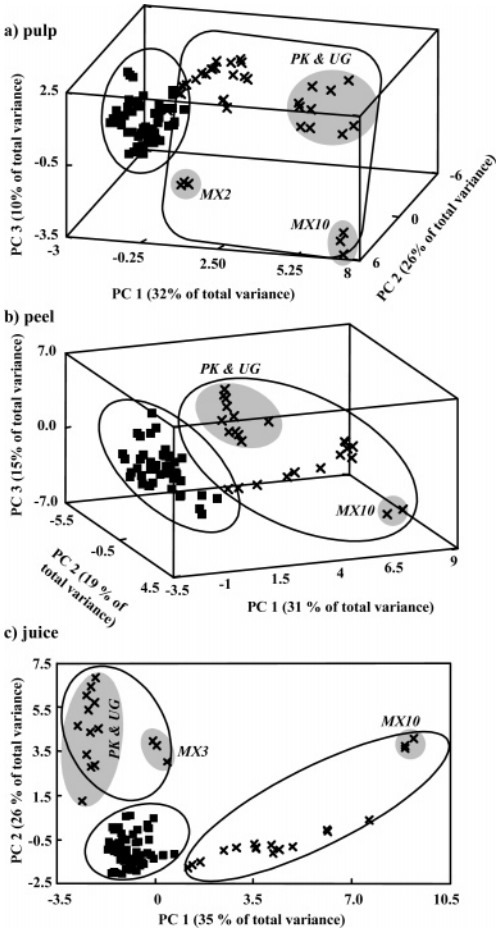

Figure 3. Projection of apple samples on the multidimensional space defined by the principal components: (**a**) pulp, (**b**) peel, and (**c**) juice. Sample codes: ■, nonbitter; ×, bitter.

major contribution to PC2 is because of dihydrochalcones [PLG, hydroxyphloretin monoglycoside (PLD-2), and PLXG] and hydroxycinnamic acids (PCQ and CAA-1) (**Table 6**). It is interesting to point out that in the three apple materials (pulps, peels, and juices), the major influential features in the PC1 were the flavan-3-ols and the dihydrochalcones; so, these polyphenol classes are those that mostly contribute to the differentiation between both categories (bitter and nonbitter). In this sense, bitter apple cultivars present higher concentrations of flavan-3-ols and/or dihydrochalcones than nonbitter varieties.

Supervised Pattern Recognition Methods. In KNN, the number of neighbors (K) assayed, in the preliminary study using a training set with all of the samples, were three, five, seven, and nine. In apple pulps and peels, none of the samples was misclassified; therefore, $K = 5$ was selected. Instead, with apple juice data, for $K \geq 5$, one (bitter) sample was wrongly classified by the model; thus, $K = 3$ was chosen.

Regarding the neural network, some empirical preliminary trials were performed to determine an adequate MLF-ANN structure. The neural architecture, which gave better results, was a MLF-ANN with three layers: an input layer with 18, 27, or 19 neurons for pulps, peels, or juices, respectively; one hidden layer with three neurons; and an output layer consisting of a neuron with a binary output (**Table 7**). Classification results afforded with each multivariate technique for each apple material are shown in **Table 8**.

Apple Pulp. Excellent results were obtained by KNN since the recognition and prediction abilities were 100% for both categories; so, all samples were correctly classified. LDA and SIMCA also achieved such good results for the nonbitter and the bitter categories, respectively. In contrast, the prediction ability of LDA for bitter cultivars was 90.7%. SIMCA was the technique that provided the worst results for classifying nonbitter varieties, showing recognition and prediction abilities of 90.3 and 76.8%, respectively. The sensitivity and the specificity of the nonbitter SIMCA model were 74 and 100%, respectively, so this model recognized 74% of nonbitter samples and rejected all of the bitter cultivars. The bitter model presented a sensitivity of 72% and a specificity of 99.6%; that is, it accepted as bitter 72% of bitter samples and 0.4% of nonbitter ones. Both models were very selective since the nonbitter model did not classify any bitter cultivar as nonbitter, and with the bitter model, only 0.4% of the nonbitter varieties could be assigned to the bitter category. Coomans plot for the squared SIMCA distances, obtained from the complete data set, allowed us to visualize SIMCA results (**Figure 4**). The different results achieved by SIMCA with regard to the other multivariate techniques used could be explained by the fact that SIMCA is a disjoint class modeling technique; therefore, more emphasis was placed on a similarity within a class than on discrimination between classes. Thus, SIMCA creates a "hyperbox" of confidence for each class: The greater the class variability is, the bigger the box will be, being able to overlap with the box of the other class. This happened with the bitter model, which presented an inappropriate specificity. The PLS-1 model constructed with apple pulp data consisted of four latent variables that explained 90.7% of the binary response variance of the cross-validation, showing a multiple linear correlation coefficient of 95%. Satisfactory results were achieved by this model with success in recognition and prediction of 100%. Box and whiskers plot for the established categories and estimated values by the PLS model for the criterion variable demonstrated that an excellent differentiation between bitter and nonbitter cultivars was afforded (**Figure 5**). MLF-ANN results were promising since it presented prediction abilities of 99.2 and 94.6% for nonbitter and bitter categories, respectively. Classical chemometric techniques for multivariate data analysis used were complementary, since LDA detected 100% of nonbitter samples, whereas SIMCA afforded a model for the nonbitter class that rejected 100% of bitter varieties. On the other hand, KNN and PLS permitted the correct classification of the 100% of bitter and nonbitter samples.

Apple Peel. LDA results were comparable for both categories; thus, it recognized correctly 100% of the samples, whereas prediction abilities were not so satisfactory, presenting percentages lower than 90%. KNN attained recognition and prediction abilities of 100% for the nonbitter category and higher than 95% for the bitter. As in pulp, SIMCA classified correctly all bitter varieties, whereas the results for the nonbitter class were considerably worse, showing recognition and prediction abilities of 85.3 and 76.5%, respectively. Sensitivity and specificity were estimated for each model established by SIMCA, being 71 and 100% for the nonbitter category and 50 and 98.9% for the bitter one, respectively. The interpretation of these results revealed that the nonbitter model was very selective since it rejected all bitter samples and accepted a high percentage (71%) of nonbitter. On the other hand, the bitter model, even though it rejected a high percentage of nonbitter samples (98.9%), recognized only 50% of the bitter samples. As it was mentioned for pulp, these results were due to the great variability of the polyphenolic profiles of the cultivars that composed the bitter category. Coomans plot in **Figure 4** shows SIMCA results. In

Table 6. Loadings of the Three PCs (PC1, PC2, and PC3) for Each Apple Material

	pulp			peel			juice		
variable	PC1	PC2	PC3	PC1	PC2	PC3	PC1	PC2	PC3
CAT	0.317	0.086	–0.296	0.282	0.135	–0.112	0.194	0.285	–0.189
EC	0.317	0.262	–0.171	0.303	0.007	–0.173	0.342	0.175	–0.081
PB2	0.250	0.339	–0.104	0.275	–0.129	–0.123	0.354	0.086	–0.076
CAT-2	0.254	0.335	–0.115	0.296	–0.093	–0.120	0.354	0.027	–0.22
PC	0.265	0.319	–0.009	0.225	–0.201	–0.169	0.354	0.083	–0.105
CA	0.103	0.158	0.435	0.143	–0.153	0.006	0.210	0.019	0.421
CAA-1	0.253	–0.293	0.265	0.192	–0.019	0.098	–0.156	0.306	–0.192
CMA-2	0.155	–0.277	0.005	0.199	0.211	0.205	–0.155	0.24	–0.366
CAA-2				0.030	0.032	–0.311	0.222	0.106	0.272
PCQ	0.264	–0.284	0.218	0.257	0.100	0.149	–0.108	0.401	0.059
PLD-1	0.248	–0.268	–0.209	0.194	0.079	–0.154	–0.167	0.279	0.359
PLD-2	0.243	–0.285	–0.176	0.191	0.158	–0.029	–0.132	0.384	0.044
PLXG	0.300	–0.054	0.006	0.298	0.023	–0.106	0.114	0.363	0.107
PLG	0.308	–0.254	–0.053	0.271	0.187	0.055	–0.094	0.416	–0.037
HYP	–0.029	0.038	0.082	0.077	–0.325	0.251	0.111	0.021	0.371
IQC	0.154	0.136	0.429	0.123	–0.306	0.109	0.254	0.072	0.144
QG-1	0.249	–0.069	0.380	0.132	–0.346	0.178	0.059	0.036	–0.313
QG-2				–0.04	–0.218	0.29			
QG-3				–0.071	–0.115	0.239			
AVI				0.212	–0.239	0.163			
QCI	0.173	0.273	0.125	0.159	–0.338	0.051	0.350	0.029	0.062
IDE				0.152	0.292	0.26			
CG-1				0.138	0.294	0.274			
CG-2				0.172	–0.003	–0.281			
CG-3				0.073	0.144	0.304			
CG-4				0.048	0.081	0.285			
DPn	–0.101	0.088	0.352	–0.138	–0.138	–0.132	0.213	–0.104	–0.252

Table 7. MLF-ANN Architectures Assayed and Their Prediction Abilities for Bitter and Nonbitter Apple Cultivars

apple material	MLF-ANN architecture	prediction ability (%)	RMSE
pulp	18, 3, 1	97.4	0.04
	18, 5, 1	93.2	0.07
	18, 3, 2	95.8	0.05
peel	27, 3, 1	96.5	0.03
	27, 5, 1	91.7	0.06
	27, 3, 2	96.5	0.05
juice	19, 3, 1	96.9	0.06
	19, 5, 1	96.4	0.04
	19, 7, 1	95.8	0.04
	19, 3, 2	94.8	0.06

PLS-1 analysis, UG variety (2000 season) was removed from the PLS-1 model, since it presented a relatively high residual error. The PLS-1 model built consisted of one latent variable that explained 82.7% of the variance of the binary response in the cross-validation, exhibiting a multiple correlation coefficient of 92%. Recognition and prediction ability were of 100% for the nonbitter class, and of 100 and 90.9%, respectively, for the bitter class. A box and whiskers plot of the predicted values by the model for the established categories allowed us to visualize that the differentiation between both categories was adequate (**Figure 5**). KNN or PLS and SIMCA were complementary for achieving a technological classification system of apple cultivars (bitter/nonbitter), attaining a level of hits of 100%. In this sense, KNN or PLS detected all nonbitter varieties, and SIMCA provided a nonbitter model that excluded all bitter varieties.

Apple Juice. LDA achieved a level of correct classification of 100% for both bitter and nonbitter samples. KNN was also effective for both categories, showing recognition and prediction abilities for the bitter category slightly lower than for the nonbitter category but higher than 95%. Regarding the SIMCA fundamentals mentioned above, some observations from the bidimensional plot of the samples on the two PCs obtained by PCA were considered (**Figure 3**). Thus, this plot revealed that samples were grouped in three regions: two groups, with bitter cultivars, and the other, with nonbitter. Hence, to perform SIMCA, three classes were established as follows: class 1, nonbitter varieties; class 2, bitter varieties without PK, UG, and MX3; and class 3, the bitter varieties PK, UG, and MX3. Classifications made by SIMCA models were notably satisfactory for the three classes, with success in recognition and prediction abilities between 95 and 100%. Sensitivities of the three models were similar, about 78%. Specificities were 100% for class 1 with regard to classes 2 and 3; for class 3 in relation to classes 1 and 2; and for class 2 related to class 3. Specificity of the model for class 2 with regard to class 1 was 95%. Thus, the three models recognized about 78% of their samples and rejected all of the samples belonging to other classes, except for the model of class 2, which accepted 5% of nonbitter samples. These results pointed out that all of the models were very selective, especially those of classes 1 and 3. In **Figure 4**, a Coomans plot represents these SIMCA results. The PLS-1 model constructed was composed by five latent variables that explained 91.5% of the binary response variance in the cross-validation, GG variety (2001 season) and a UG sample (2000 season) having been removed from the model because they were leverage points. The multiple correlation coefficient was 96%, and their recognition and prediction capacities were 100% for both categories, allowing a correct differentiation between them. The box and whiskers plots of these results are shown in **Figure**

Table 8. Classification Results for the Supervised Pattern Recognition Techniques Applied to Apple Data

		pulp		peel		juice	
technique	class	recognition ability (%)	prediction ability (%)	recognition ability (%)	prediction ability (%)	recognition ability (%)	prediction ability (%)
LDA	nonbitter	100.0	100.0	100.0	88.6	100.0	100.0
	bitter	100.0	90.7	100.0	84.6	100.0	100.0
KNN; inverse squared Euclidean distance[a]	nonbitter	100.0	100.0	100.0	100.0	100.0	100.0
	bitter	100.0	100.0	98.7	95.8	99.5	95.3
PLS-1	nonbitter	100.0	100.0	100.0	100.0	100.0	100.0
	bitter	100.0	100.0	100.0	90.9	100.0	100.0
MLF-ANN; $\eta = 0.2$; $\mu = 0.5$; sigmoidal transfer function[b]	nonbitter	100.0	99.2	100.0	96.9	100.0	98.4
	bitter	100.0	94.6	100.0	95.8	100.0	94.0
SIMCA; normal range; $\alpha = 0.05$[c]	1 (nonbitter)[d]	90.3	76.8	85.3	76.5	94.8	98.9
	2 (bitter)[e]	100.0	100.0	100.0	100.0	100.0	100.0
	3 (bitter)[f]					100.0	100.0

[a] K = five in pulp and peel; K = three in juice. [b] MLF-ANN architecture: $(18 \times 3 \times 1)$ in pulp; $(27 \times 3 \times 1)$ in peel; and $(19 \times 3 \times 1)$ in juice. [c] Six PCs for each category in pulp and peel and three PCs for each catergory in juice. [d] In juice, class 1: nonbitter cultivars. [e] In juice, class 2: bitter cultivars without PK, UG, and MX3. [f] In juice, class 3: bitter cultivars PK, UG, and MX3.

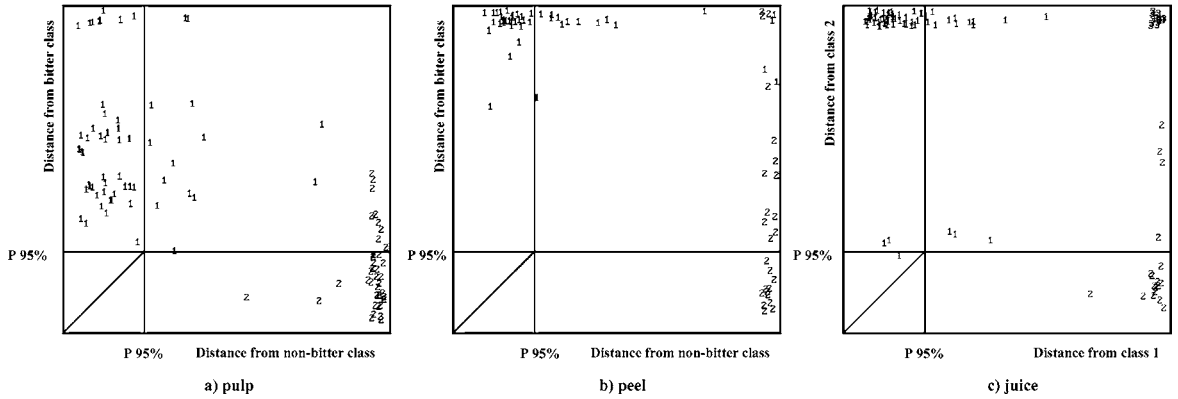

Figure 4. Coomans plot for the squared SIMCA distances for apple data. Sample codes for (**a**) pulp and (**b**) peel: 1, nonbitter; 2, bitter. Sample codes for (**c**) juice: 1, class 1 (nonbitter); 2, class 2 (bitter without PK, UG, and MX3); 3, class 3 (bitter PK, UG, and MX3).

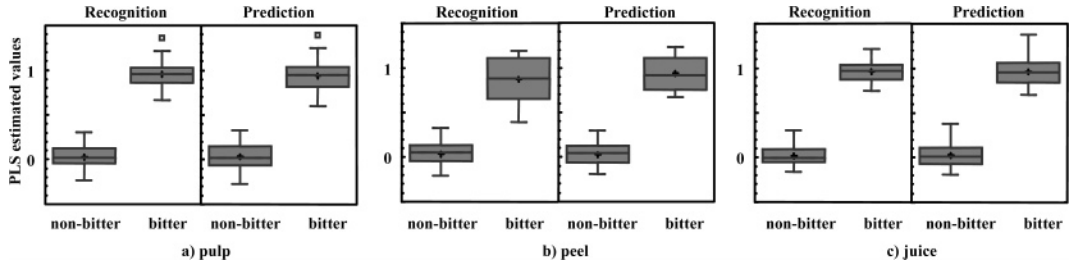

Figure 5. Multiple box and whisker plots for PLS estimated values: (**a**) pulp, (**b**) peel, and (**c**) juice.

5. With apple juice data, KNN and SIMCA afforded complementary models, since KNN identified every nonbitter cultivar, and SIMCA proposed a model for the nonbitter class that rejected all bitter varieties.

Apple peel composition depends to a great extent on climatology (*7*), sun exposition of the fruit—existing differences depending on the position of the fruit in the tree and even, in the same fruit, between sun-exposed parts and shaded parts (*33*). On the other hand, juice-making procedures influence the polyphenolic composition of juices. Therefore, bitterness classification made with pulp data was considered to be more accurate. Using the classification systems obtained, the bitterness of all cider apple cultivars studied was established (**Table 4**).

In conclusion, bitter apple cultivars presented higher contents of flavan-3-ols and/or dihydrochalcones than nonbitter cultivars in their pulps, peels, and juices. Regarding the information given by PCA, these were the classes of polyphenols that mostly contribute to differentiate between both technological groups.

From the results obtained by the different supervised pattern recognition techniques applied, it was stated that polyphenolic profiles of apple pulps, peels, and juices contained enough and suitable information to develop classification rules to predict the bitterness of apple cultivars.

The method proposed to establish the bitterness of an apple cultivar consists of the determination of the polyphenolic profile of its pulp and the use of the classification rules developed by KNN or PLS for performing the prediction (100% of hits). The complementary LDA and SIMCA decision rules together can also achieve such good results. Thus, bitterness of all of the cider apple cultivars studied was concluded.

In addition, it was inferred that pattern recognition techniques are capable of extracting useful information from a huge data set, to relate the chemical composition of cider apple pulps, peels, and juices with their sensory and technological properties.

ABBREVIATIONS USED

AVI, avicularin; PB2, procyanidin B2; CQA, 5-caffeoylquinic acid; CAA-1, -2, unknown hydroxycinnamic acids with caffeic acid UV spectra; CAT, (+)-catechin; CAT-2, unknown flavan-3-ol; CG-1, -2, -3, -4, unknown anthocyanins; CMA-2, unknown hydroxycinnamic acid with *p*-coumaric acid UV spectra; DPn, average degree of polymerization of procyanidins; EC, (−)-epicatechin; HYP, hyperoside; IDE, ideain; IQC, isoquercitrin; PC, total procyanidins; PCQ, 4-*p*-coumaroylquinic acid; PLD-1, hydroxyphloretin diglycoside; PLD-2, hydroxyphloretin monoglycoside; PLG, phloridzin; PLXG, phloretin-2′-O-xyloglucoside; PPO, polyphenoloxidase; QCI, quercitrin; QG-1, -2, -3, unknown flavonols; RUT, rutin; CA, cluster analysis; KNN, K-nearest neighbors; LDA, linear discriminant analysis; MLF-ANN, multilayer feed forward artificial neural network; PCA, principal component analysis; PCs, principal components; PC1, first principal component; PC2, second principal component; PC3, third principal component; PLS, partial least-squares; PRMSE, prediction root mean square error; RMSE, root mean square error; SD, standard deviation; SIMCA, soft independent modeling of class analogy; DAD, diode array detector; HPLC, high-performance liquid chromatography; ND, not detected; t, traces; AG, Azpuru Garratza; BK, Bost Kantoi; ER, Errezila; GG, Gazigorri; GK, Goikoetxea; GM, Geza Miña; GZ, Gazilokia; IB, Ibarra; LR, Larrabetzu; MK, Moko; MN111, Manttoni 111; MNEM7, Manttoni EM7; MX1, Mendexa 1; MX10, Mendexa 10; MX11, Mendexa 11; MX2, Mendexa 3; MX3, Mendexa 2; MX4, Mendexa 4; MZ, Mozoloa; PK, Piko; PL, Palazio; PT, Patzuloa; TT, Txistu; TX, Txalaka; UG, Ugarte; UGS, Urdai Goika Santutxu; UH, Urtebi Haundia; UM, Udare Marroi; UR, Urdin; URZ, Urdin Zalla; UT, Urtebi Txiki.

ACKNOWLEDGMENT

We express our gratitude to the Diputación Foral de Gipuzkoa and the Diputación Foral de Bizkaia for providing apple samples and to Dr. S. Guyot (INRA, France), Dr. F. A. Tomás-Barberán (CEBAS-CSIC, Spain) and Dr. C. Santos-Buelga (Universidad de Salamanca, Spain) for providing polyphenol standards.

LITERATURE CITED

(1) Cliff, M.; Dever, M. Characterization of varietal apple juices. *Can. Inst. Food Sci. Technol. J.* **1990**, *23*, 217−222.

(2) Blanco-Gomis, D.; Herrero-Sánchez, I.; Mangas, J. J. Characterization of apple cider cultivars by chemometric techniques using data from high-perfomance liquid chromatography and flow-injection analysis. *Analyst* **1998**, *123*, 1187−1191.

(3) Dapena, E. Comportamiento agronómico y tecnológico de variedades de manzano asturianas. Ph.D. Thesis, Universidad de Oviedo, Oviedo, Spain, 1996.

(4) Sanoner, P.; Guyot, S.; Marnet, N.; Molle, D.; Drilleau, J. F. Polyphenol profiles of French cider apple varieties (Malus domestica sp.). *J. Agric. Food Chem.* **1999**, *47*, 4847−4853.

(5) Lea, A. G. H. Cidermaking. In *Fermented Beverage Production*; Lea, A. G. H., Piggot, J. R., Eds.; Blackie Academic & Professional: London, U.K., 1995; pp 66−96.

(6) Cowan, M. M. Plant products as anti-microbial agents. *Clin. Microbiol. Rev.* **1999**, *12*, 564−582.

(7) Lea, A. G. H. Bitterness and astrigency: The procyanidins of fermented apple ciders. In *Bitterness in Food and Beverages*; Rousself, R. L., Ed.; Elsevier: Oxford, U.K., 1990; pp 123−143.

(8) Lea, A. G. H. Flavor, color, and stability in fruit products: The effect of polyphenols. In *Plant Polyphenols. Synthesis, Properties, Significance*; Hemingway, R. W., Laks, P. E., Eds.; Plenum Press: London, U.K., 1992; pp 827−848.

(9) Amiot, M. J.; Tacchini, M.; Aubert, S.; Nicolas, J. Phenolic composition and browning susceptibility of various apple cultivars at maturity. *J. Food Sci.* **1992**, *57*, 958−962.

(10) Lea, A. G. H.; Arnold, G. M. The phenolics of ciders: Bitterness and astringency. *J. Sci. Food Agric.* **1978**, *29*, 478−483.

(11) Pérez-Ilzarbe, J.; Hernández, T.; Estrella, I. Phenolic compounds in apples: Varietal differences. *Z. Lebensm. Unters. Forsch.* **1991**, *192*, 551−554.

(12) Escarpa, A.; González, M. C. High-performance liquid chromatography with diode-array detection for the determination of phenolic compounds in peel and pulp from different apple varieties. *J. Chromatogr. A* **1998**, *823*, 331−337.

(13) Mangas, J. J.; Rodríguez, R.; Suarez, B.; Picinelli, A.; Dapena, E. Study of the phenolic profile of cider apple cultivars at maturity by multivariate techniques. *J. Agric. Food Chem.* **1999**, *47*, 4046−4052.

(14) Alonso-Salces, R. M.; Korta, E.; Barranco, A.; Berrueta, L. A.; Gallo, B.; Vicente, F. Determination of polyphenolic profiles of Basque cider apple varieties using accelerated solvent extraction. *J. Agric. Food Chem.* **2001**, *49*, 3761−3767.

(15) Price, K. R.; Prosser, T.; Richetin, A. M. F.; Rhodes, M. J. C. A comparison of the flavonol content and composition in dessert, cooking and cider-making apples; distribution within the fruit and effect of juicing. *Food Chem.* **1999**, *66*, 489−494.

(16) Planton, G. *Le point sur le test almidon des pommes pour l'aide á la décision de récolte*; Ctifl: France, 1995; no. 6.

(17) Guyot, S.; Marnet, N.; Sanoner, P.; Drilleau, J. F. Direct thiolysis on crude apple materials for the HPLC characterization and quantification of polyphenols in cider apple tissues and juices. *Methods Enzymol.* **2001**, *335*, 57−70.

(18) Alonso-Salces, R. M.; Barranco, A.; Corta, E.; Berrueta, L. A.; Gallo, B.; Vicente, F. A validated solid−liquid extraction method for the HPLC determination of polyphenols in apple tissues. Comparison with pressurised liquid extraction. *Talanta*, in press.

(19) Singleton, V. L.; Rossi, J. A. Colorimetry of total phenolics with phosphomolybdic-phosphotungstic acid reagents. *Am. J. Enol. Vitic.* **1965**, *16*, 144−158.

(20) Cazallas, R.; Fernández, L. A.; Etxebarria, N.; Madariaga, J. M. Automated system for multiple potenciometric and spectrophotometric titrations. *LRA* **1993**, *5*, 161−169.

(21) Cazallas, R. Diseño de sistemas automáticos de medidas potenciométricas y conductimétricas. Aplicación en el control de procesos químicos y en el estudio de equilibrios químicos en disolución. Ph.D. Thesis, Universidad del País Vasco, Leioa, Spain, 1997.

(22) Padín, P. M.; Peña, R. M.; García, S.; Iglesias, R.; Barro, S.; Herrero, C. Characterization of Galician (N. W. Spain) quality brand potatoes: A comparison study of several pattern recognition techniques. *Analyst* **2001**, *126*, 97−103.

(23) LaTorre, M. J.; Peña, R.; García, S.; Herrero, C. Authentification of Galician (N. W. Spain) honeys by multivariate techniques based on metal content data. *Analyst* **2000**, *125*, 307−312.

(24) *Statgraphics Plus 5.0*; Statistical Graphics Corp.: Rockville, MD, 1994−2000.

(25) Forina, M.; Lanteri, S.; Armanino, C. *Q-Parvus 3.0. An Extendable Package of Programs for Data Explorative Analysis, Classification and Regression Analysis*; Dipartimento Chimica e Tecnologie Farmaceutiche ed Alimentari: University of Genova, Genova, Italy, 2000.

(26) *SPSS for Windows*, version 9.0; SPSS Inc.: Chicago, IL, 1989−1999.

(27) *The Unscrambler*, version 7.5; CAMO ASA: Woodbridge, NJ, 1986−1999.

(28) Massart, L.; Kaufman, L. *The Interpretation of Analytical Chemical Data by the Use of Clusters Analysis*; Wiley-Interscience: New York, 1983.

(29) Meléndez, M.; Ortíz, M. C.; Sánchez, M.; Sarabia, L.; Iñiguez, M. Chemometric characterization of the clarets and rose wines of the certified denomination of origin Rioja using Cielab parameters. *Quím. Anal.* **1999**, *18*, 119−126.
(30) *WinNN32 1.2*; Y. Danon: Arad, Israel, 1993−1996.
(31) Alonso-Salces, R. M.; Barranco, A.; Abad, B.; Berrueta, L. A.; Gallo, B.; Vicente, F. Polyphenolic profiles of Basque cider apple cultivars and their technological properties. *J. Agric. Food Chem.* **2004**, *52*, 2938−2952.
(32) Janovitz-Klapp, A. H. Etudes sur le système polyphenoloxidase isolé de la pomme (*Malus sylvestris*, Var. Red Delicious). Ph.D. Thesis, Université de Paris, Paris, France, 1989.
(33) Awad, M. A.; De Jager, A.; Van Westing, L. M. Flavonoid and chlorogenic acid levels in apple fruit: Characterisation of variation. *Sci. Hortic.* **2000**, *83*, 249−263.

Received for review June 21, 2004. Revised manuscript received October 11, 2004. Accepted October 14, 2004. This research was supported by Gobierno Vasco/Eusko Jaurlaritza (Projects PI-1997-19 and PI-1999-106) and Universidad del País Vasco/Euskal Herriko Unibertsitatea (Project 171.310-EB013/98). R.M.A.-S. thanks Gobierno Vasco/Eusko Jaurlaritza for a Ph.D. grant.

JF048995I