

Characterization of Volatile Compounds in New Cider Apple Genotypes Using Multivariate Analysis

Jairo Pello-Palma¹ · Juan José Mangas-Alonso² · Enrique Dapena de la Fuente² · Jaime González-Álvarez¹ · Jorge Díez³ · María Dolores Gutiérrez Álvarez¹ · Pilar Arias Abrodo¹

Received: 26 November 2015 / Accepted: 29 April 2016
© Springer Science+Business Media New York 2016

Abstract Gas chromatography combined with solid-phase microextraction has been used for the identification of the aromatic profiles of new cider apple genotypes, and a chemometric characterization of these new cider apple genotypes has been carried out using exploratory and modelling techniques. Three breeding targets have been explored: (1) regular bearing and scab resistance, (2) resistance to bio-aggressors and (3) high polyphenol content and late ripening. Exploratory techniques established two genotype groups: those that come from breeding towards targets 1 and 2 with low polyphenol contents and those that come from breeding towards target 3 with high polyphenol contents. Alcohols were related to the genotypes with breeding towards target 3, and compounds such as esters were related to the genotypes with breeding towards targets 1 and 2. Models computed using the soft independent modelling of class analogy (SIMCA) technique presented good sensitivity (93 %), specificity (91 %) and classification hits (96 %). However, the predictions computed by SIMCA (70 %) and the artificial neural network (ANN) (76 %) were low.

Electronic supplementary material The online version of this article (doi:10.1007/s12161-016-0521-7) contains supplementary material, which is available to authorized users.

✉ Jaime González-Álvarez
gonzalezalvarezjaime@gmail.com

✉ Pilar Arias Abrodo
piarab@uniovi.es

¹ Department of Physical and Analytical Chemistry, University of Oviedo, Julián Clavería 8, 33006 Oviedo, Spain

² Servicio Regional de Investigación y Desarrollo Agroalimentario (SERIDA), Villaviciosa, Spain

³ Artificial Intelligence Center, University of Oviedo, Campus de Viesques, Gijón 33204, Spain

Keywords Cider apple · Breeding · Chemometric · HS-GC · SPME · Volatile compounds

Introduction

Apples are one of the most consumed fruits across the world and are mainly grown in temperate regions. The number of apple varieties that have been documented is enormous (more than 10,000), resulting in a wide range of properties such as resistance, sweetness, acidity and ripening. However, currently, only a few types of apples dominate world production (Janick et al. 1996). In Asturias, a region in northwest Spain, there is an important production of cider apples (more than 35,000 t per year), which are mainly used to make cider and other derivatives. Furthermore, there are a large number of local cider apple cultivars, and this region could be considered as an area with a great genetic diversity of apples.

There is a current interest in the development of new varieties of apples, but the requirements are constantly changing and becoming more challenging. The tasks concern issues such as the rates and character of tree growth, resistance to different diseases, period of maturation, productivity, commercial and consumer appeal of the fruit and agro-technical methods (Bus et al. 2012; Troggo et al. 2012; Kumar et al. 2010). To obtain new cider apples to improve the cider quality of the protected designation of origin (PDO) ‘Cider of Asturias’ and other derivatives as juice of cider apple, a cider apple breeding program has been carried out in Asturias beginning in 1989 (Dapena and Blázquez 2004). In this program, three lines of crossings were performed in order to achieve cultivars with the following characteristics: (1) regular bearing and resistance to scab, (2) resistance to some bio-aggressors (scab and fire blight pathogens as well as rosy apple aphid) and (3) late maturity apples with a high content

of phenols. Fifty-four new genotypes of this program were preselected, and a deep knowledge about their volatile composition will be of great interest for the final selection process of these genotypes.

The volatile compounds have shown a great importance on the sensorial quality of fresh and processed fruit products (Ulrich et al. 2009; Fan et al. 2011; Kafkas et al. 2013; Schwab et al. 2008). While the role of sugars and acids in the taste are well known and have been described, the participation of the volatile compounds towards the flavour and odour is more difficult to define. The contribution of the aroma to the flavour quality of fresh product has attracted increasing attention, and breeders need more information and analytical tools in order to select flavour quality (Baldwin 2004). A large number of volatile compounds have been measured in the aroma profiles of apples, but only a few of them seem to dominate the typical aroma of apples (Dixon and Hewett 2000; Fuhrmann and Grosch 2002; Dunemann et al. 2009). The following three main groups of compounds can be identified in the flavours of apples: esters, aldehydes and alcohols, with esters being the most important (Kafkas et al. 2013).

In the few last years, some works have been conducted on the analysis of the volatile compounds in the progeny of programs to improve apple dessert varieties (Rowan et al. 2009a, b) and to identify the quantitative trait locus (QTL) associated with volatile organic compounds (Zini et al. 2005; Dunemann et al. 2009; Costa et al. 2013).

In contrast, only a limited number of studies have been addressed towards the identification of the volatile profiles of cider apples, in particular the PDO Asturian cider apple (Arias Abrodo et al. 2010; Díaz Llorente et al. 2010; Díaz-Llorente et al. 2011; Díaz-Llorente et al. 2012).

Currently, the most common analytical techniques used to determine apple aromatic profiles are chromatographic, in particular gas chromatography (GC). The complexity of fruit matrices requires the isolation of the analytes as precursors before their introduction into the chromatographic system. To achieve this aim, a variety of separation techniques (Sampaio et al. 2014) have been proposed, including solid-phase microextraction, which has been widely employed in the sample preparation (Pawliszyn 2012). The main advantages of this technique are the simplicity of operation, speed, absence of solvents and preconcentration.

To take into account the multivariate nature of the data, multivariate chemometric data analysis techniques have to be applied (Daszykowski et al. 2007).

Analysis techniques such as exploratory (cluster analysis, self-organizing maps (SOMs)), factorial and principal component are used to visualize the structure of the database and for searching for the relationship between features and examples (Blanco-Gomis et al. 2009; Wehrens and Buydens 2007). Multivariate classification for qualitative analysis by means of supervised pattern recognition methods, such as linear

discriminant analysis (Ballabio and Todeschini 2009), Bayesian analysis (quadratic discriminant analysis) (Hibbert and Armstrong 2009), soft independent modelling of class analogy (SIMCA) (Fernández et al. 2007), multivariate regression (partial least squares (PLS)) (Cozzolino et al. 2009) or artificial neural network (ANN) (Chen et al. 2009), was carried out.

The main goal of this study has been to characterize the new cider apple genotypes obtained from the Asturias breeding program by multivariate analysis of their volatile compound profile in order to contribute to the selection and inclusion of those that present higher interest for incorporation into the PDO Cider of Asturias and for the production of high-quality apple derivatives.

Experimental

Reagents and Materials

Individual volatile compounds with purities over 99 % were supplied by Sigma-Aldrich (Madrid, Spain). Methanol (HPLC grade) and sodium chloride were obtained from Merck (Darmstadt, Germany). Milli-Q ultrapure water (Millipore, Bedford, MA, USA) was used throughout the analyses.

Glass vials (20 mL) with PTFE septa and a manual solid-phase microextraction (SPME) holder, 65 μm polydimethylsiloxane-divinylbenzene (PDMS/DVB), 100 μm polydimethylsiloxane (PDMS), 75 μm carboxen-polydimethylsiloxane (CAR/PDMS) and 50/30 μm divinylbenzene-carboxen-polydimethylsiloxane (DVB/CAR/PDMS) fibres were obtained from Supelco (Bellefonte, PA, USA).

A stock standard solution of volatile compounds was prepared at a concentration of 1000 mg L^{-1} by diluting the appropriate volume of the pure compound in methanol/water (50/50) solvent. The standard solution was stored at 4 °C and protected from light. Working solutions were prepared by dilution in water of the appropriate volume of the above solution.

Sample Preparation

Genotypes that belong to the three breeding lines mentioned above (namely 1, 2 and 3) were studied for the 2012 and 2013 harvests. Table S1 (see Supporting Information) shows the 54 genotypes analysed showing information about the crossing line, parental genitors and technological class of each genotype. A sample of 50 pieces of each apple genotype was picked from the tree at the optimal harvest time by using the starch-iodine test. This maturity index was used in agreement

with Díaz-Llorente et al. (2012) that established a mathematical model to predict this index from nine apple aromas.

Three batches of 1–2 kg of apples per genotype were milled and pressed in a small hydraulic press (Hafico). Afterwards, the apple juice obtained was clarified by centrifugation (8000 rpm, 10 min, 10 °C). The final apple juice samples were kept at –20 °C until the analyses were performed.

Identification of the Volatile Compounds by GC-MS

The volatile compounds of the apple juice samples were identified using an Agilent 6890 GC system coupled with an Agilent MD5973 quadrupole mass spectrometer. The separation of the extracted compounds was performed using a polyethylene glycol-acidified capillary column (50 m × 0.25 mm I.D. × 0.25 µm film thickness) purchased from Teknokroma (Barcelona, Spain).

All SPME extractions were conducted with a 65 µm PDMS/DVB fibre. Desorption of the fibre into the injection port was carried out in the splitless mode at 250 °C for 2 min. The program temperature used for the chromatographic separations was as follows: An initial temperature of 60 °C was held for 1 min and increased to 220 °C at 5 °C min⁻¹ where it was then held for 1 min.

The mass spectrometer was operated in scan mode from $m/z = 33$ to 300 at 10 scans/s, and the ionization was performed with a kinetic energy of the impacting electrons of 70 eV.

Twenty-nine parental genotypes and local cider apple cultivars were studied, and the following volatile compounds were selected for the characterization of different cider apple genotypes: ethyl acetate, ethyl propanoate, ethyl butyrate, ethyl 2-methylbutyrate, butyl acetate, isopentyl acetate, butyl butyrate, butyl 2-methylbutyrate, hexyl acetate, hexyl 2-methylbutyrate, 1-butanol, 2-methyl-1-butanol, 1-pentanol, 1-hexanol, trans-2-hexen-1-ol, hexanal and 6-methyl-5-hepten-2-one. These volatile compounds were selected according to their relative abundance and their impact on the aroma profile.

Identification of the volatile compounds was carried out by using the retention indices of reference compounds and by comparison between the mass spectra of the unknown compounds, and the mass spectra of standard compounds compiled in a Wiley 275 library.

High-Speed Gas Chromatography Conditions

Analyses were performed on a Shimadzu GC2010 high-speed gas chromatograph (Shimadzu, Kyoto, Japan) with flame ionization detection (FID). The GC was equipped with a TRB-FFAP capillary column (10 m × 0.1 mm I.D. × 0.20 µm film thickness).

Helium with a flow rate of 0.49 mL min⁻¹ was used as the carrier gas. The GC was operated in the splitless mode, and the

injector and detector temperatures were 250 and 280 °C, respectively. The temperature was programmed as follows: An initial temperature of 35 °C for 4 min increased at a rate of 10 °C min⁻¹ to 70 °C followed by holding for 1 min, increased at a rate of 15 °C min⁻¹ to 130 °C followed by holding for 0.01 min, and increased to 200 °C at a ramp of 70 °C min⁻¹.

To illustrate the results obtained, chromatograms of volatile compounds of each of the breeding lines studied herein are shown in Fig. S1 (see Supporting Information).

Experimental Design

A 2³ full factorial design (FFD) was performed to investigate the effects of the extraction time, temperature and equilibration time in order to obtain the maximum response in the aroma extraction process. Low and high levels were selected based on preliminary tests. The order of the experiments was fully randomized to avoid possible memory effects of the analytical apparatus. This experimental plan allows the evaluation of the effects of the main factors and their interactions. In this case, NaCl (4 g) and the stirring speed (600 rpm) were fixed. In all cases, data analyses were performed by means of the statistical package Statgraphics Centurion XV for Windows version 15.2.06 manufactured by Statpoint Technologies (Warrenton, VA, USA).

Data Analysis

The data matrix consisted of 17 columns (aroma compounds) and 90 rows (cider apple genotypes), and it was categorized as follows: class 1 (cases of breeding line 1 (harvest years of 2012 and 2013), $N=40$), class 2 (cases of breeding line 2 (harvest years of 2012 and 2013), $N=34$) and class 3 (cases of breeding line 3 (harvest years of 2012 and 2013), $N=16$) (see Table S2 in Supporting Information). Before the multivariate analysis, the data were autoscaled and cluster analysis was used in order to compute the level of similarity between the samples belonging to each batch. When the level of similarity was higher than 95 %, the average of the three batches for each aroma was computed before multivariate analysis. In order to select variables, a similarity matrix of them by means of Euclidean metric was computed. The similarities computed were lower than 0.8 so that whole variables for multivariate analyses were used.

The data were processed using multilayer perceptron for ANN in the Weka experimental environment (Hall et al. 2009), SOM Toolbox in Matlab for SOM, Minitab for principal component analysis (PCA) and PARVUS for SIMCA and cluster analysis.

Results and Discussion

Analytical Method

Fibre Selection

Four commercially available SPME fibre coatings, with different polarities and inner structures, were tested for their efficiencies in the SPME of volatile compounds in a standard solution as mentioned above. The comparison of SPME fibre performance can be made in terms of both the extraction efficiency and the number of identifiable compounds in the extract. The PDMS-DVB fibre exhibited the highest extraction efficiency (100 %) for all of the studied analytes compared to the DVB/CAR/PDMS (62 %), CAR/PDMS (61 %) and PDMS (28 %) fibres.

Accordingly, further studies were conducted using the PDMS-DVB fibre.

Optimization of the Extraction Factors

The intervals of factors and the response variable were selected and applied from the factorial program in one block. The experimental design program proposed ten experiments with random variation of the factors for the PDMS-DVB fibre. The chosen variable was the total peak area, which was obtained from the chromatograms. After having analysed the samples in all of the conditions necessary for multivariate and predictive analysis, a chromatogram was obtained and employed for quantifying the influence of each variable. The selected intervals of equilibration and the extraction time and temperature were chosen according to two characteristics of the analytical method, namely the sensitivity and analysis time. The selected range for the equilibration time was 5 to 25 min. In this case, the chromatographic peak areas slightly increased until 25 min. Because the time required to reach this situation was considerable, we chose non-equilibrium conditions to select the range of values for the equilibration time. The extraction time was tested in a range from 5 to 15 min, and the temperature was evaluated in a range from 25 to 45 °C.

According to the results of the experimental design, an equilibration time of 5 min, a temperature of 45 °C and an extraction time of 15 min were fixed to obtain the optimal results.

Analytical Performance Characteristics

Using the appropriate optimum conditions obtained above, the analytical figures of merit were evaluated. The regression coefficients of calibration curves for all of the compounds ranged between 0.987 and 0.999, and the detection limits were from 0.001 to 0.03 $\mu\text{g L}^{-1}$. The quantification limit ranged from 0.8 to 2.7 $\mu\text{g L}^{-1}$. The instrumental repeatability ($n=6$)

checked using a 100 $\mu\text{g L}^{-1}$ solution of volatile compound standards had relative standard deviations ranging from 0.5 to 18 %, whilst the reproducibility obtained from the analysis of freshly prepared solutions on separate days was from 3 to 19 %, indicating that the proposed method is adequate for the analysis of these compounds.

The recoveries obtained (84.2–112.7 %) are in agreement with other methods using SPME.

Exploratory and Principal Component Analysis

Exploratory Data Analysis

Exploratory data analysis (EDA) allows us to find hidden and unknown information from data using graphical approach as tool. Self-organizing Kohonen's maps is a non-linear approach technique that can be used for EDA purposes. This technique was developed by Kohonen (1982), as an exploratory tool. SOM is a non-supervised ANN. The principal goal of this network is to map the examples in our dataset from a high-dimensional space to a two-dimensional one. This network can be graphically represented as a map, in which darker cells involve greater distances and where some cells present a label indicating which class is predominant in that cell.

Figure 1 shows the distribution of the classes in a two-dimensional map as well as the distances between them. As seen, the examples of class 3 are very close to those of class 1 (light cells imply closeness). There is also, at the top of the map, an area in which class 1 and class 2 are mixed, and this area is far away from the rest. Class 3 is defined by a particular breeding goal because genotypes belonging to this class were obtained by crossing a parent with a high content of phenolic compounds ($c > 2$ g tanic acid per litre, Lea and Piggot 1995) with another featuring late ripening and slight bitterness. Classes 1 and 2 contain genotypes defined by their resistance

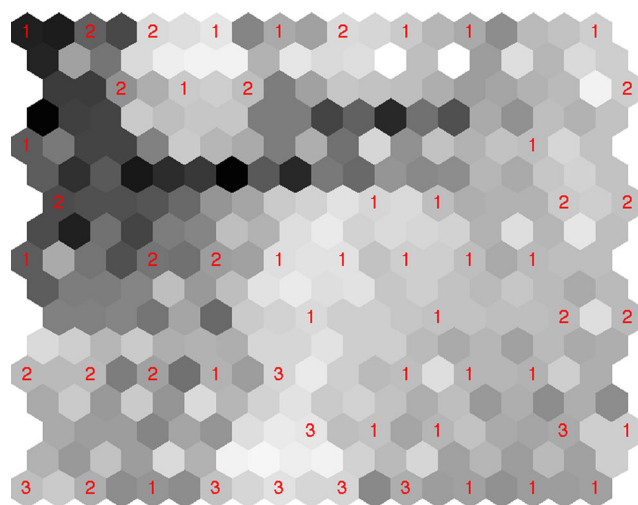


Fig. 1 Two-dimensional SOM obtained for the classes 1, 2 and 3

to scab and regular cropping (class 1). It is important to note that some genotypes belonging to class 1 could have a medium ($c > 1.5$ g tanic acid per litre, Dapena 1993) or high content of phenolic compounds, so they are placed next to class 3.

Principal Component Analysis

PCA is a bilinear decomposition/projection technique that condenses big data into few parameters named principal components. A model of seven factors that accounted for the 77 % of the variance was computed. Taking into account the information obtained from EDA, we used the first and third components of the PCA model computed in order to better visualize the data structure and the relationships between variables (volatile compounds) and observations (apple genotypes). Figure S2 (see Supporting Information) shows loadings of the first three principal components, and Fig. 2 displays the scores projected onto the plane formed by the first and third components (35 % accumulated variance). Variables with the highest and lowest loadings for the first component were 1-hexanol, 2-methyl-1-butanol, 1-butanol, three ethyl esters [propanoate, butyrate and 2-methylbutyrate], butyl acetate, hexanal, isopentyl acetate and 1-pentanol. Variables with the highest and lowest loadings for the third component were ethyl esters [acetate, propanoate, butyrate and 2-methylbutyrate], hexanal, butyl esters [butyrate and 2-methylbutyrate], 1-hexanol, hexyl acetate, 2-methyl-1-butanol and 1-butanol.

A data structure can be visualized in the score projection (Fig. 2). Genotypes derived of breeding line 3 (class 3) were placed in the lower right corner of the factorial plane. We can consider that variables with high and low loadings for the first and third principal components, such as 1-hexanol, 2-methyl-1-butanol and 1-butanol (Fig. S2), respectively, are correlated to breeding line 3 genotypes. Genotypes belonging to

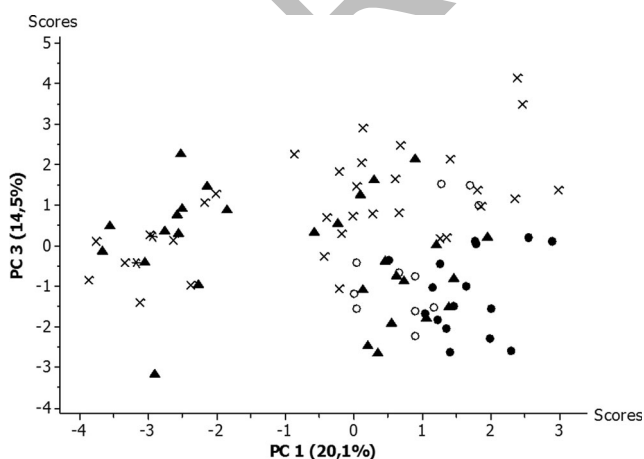


Fig. 2 Score projection onto the plane formed by the first and third principal components (breeding line 1 (circles), breeding line 2 (triangles), breeding line 3 (asterisks))

breeding lines 1 and 2 (classes 1 and 2, respectively) were not differentiated onto the factorial plane, and some genotypes belonging to breeding line 1 were projected near to the genotypes belonging to breeding line 3 (Fig. 2). It is necessary to stress that these class 1 genotypes contain high contents of phenolic compounds; therefore, their position near the class 3 genotypes might be justified because breeding line 3 has the aim of obtaining apples with high polyphenol contents. Conversely, a subgroup of genotypes belonging to classes 1 and 2 was projected on the left of the factorial plane. These apples were defined by their 1-pentanol, butyl acetate and isopentyl acetate contents. It is remarkable that these genotypes had as one of the parental varieties ‘Florina’ or ‘H232’, with both of them displaying the same aroma profile.

Modelling Techniques

Two different modelling techniques named ANN and SIMCA have been used, in order to obtain more robust results. ANN is considered a non-linear statistical data modelling tool, whilst SIMCA is a class modelling method focused on finding the similarities between samples that belong to the same category and not to find differences between two or more categories.

Artificial Neural Network

Since Rosenblatt (1958) created the perceptron, ANNs have been widely used for pattern recognition. These networks have an input layer with as many neurons as features described in each observation and an output layer with a neuron by class. It is possible to connect the input layer directly to the output layer, and in this case, we are dealing with a linear perceptron (LP). However, it is also possible to have intermediate layers (hidden layers) which lead to a multilayer perceptron (MLP). The insertion of hidden layers increases the searching space of the MLP allowing non-linear solutions. In contrast, it is more difficult to train these networks and interpret them.

LP consisted of 17 neurons in the input layer, with each one corresponding to an original variable (aroma), and three neurons in the output layer, with each one corresponding to a class defined by its breeding goal.

We used both ANN [MLP (one hidden layer with ten neurons) and LP] to try and find a model able to discriminate examples of the three classes in our dataset. In both cases, the experiment was a leave-one-out, that is, for predicting the class of one example, a model was learned using the rest of the examples as a training set.

MLP predicted correctly 75.6 % of the observations, a percentage slightly better than LP (73.3 %). Table 1 shows the confusion matrices for both ANN. Most of the examples incorrectly classified belonged to classes 2 and 3. Although MLP obtains slightly better hits than LP, the presence of a

Table 1 Confusion matrix in a leave-one-out experiment using the MLP model (top) and LP model (bottom)

MLP	Classified as	Class 1	Class 2	Class3
1 Hidden layer With 10 neurons	Class 1	30	7	3
	Class 2	9	25	0
	Class 3	3	0	13
LP	Classified as	Class 1	Class 2	Class3
	Class 1	27	10	3
	Class 2	8	26	0
	Class 3	2	1	13

hidden layer makes the identification of those features that are more relevant in order to classify the genotypes in each class very difficult. For that reason, the model computed by LP was analysed.

Table 2 shows the weights (c_{ij}) of each connection in the LP. A genotype x is assigned to class 1 if the inner product $\langle c1,x \rangle$ is higher than both $\langle c2,x \rangle$ and $\langle c3,x \rangle$. Thus, the most relevant features are those whose weights have high absolute values. As seen, objects belonging to class 1 seem to have high contents of butyl acetate ($c_{ij}=8.06$) and hexanal ($c_{ij}=7.32$) and low

Table 2 Weights learnt by the LP for each feature

	Class 1 (c1)	Class 2 (c2)	Class 3 (c3)
Variable	-0.44	0.90	-2.79
Ethyl acetate	2.30	-0.84	-0.68
Ethyl propanoate	-4.86	2.43	3.68
Ethyl butyrate	-1.83	8.60	0.34
Ethyl 2-methylbutyrate	-3.20	2.70	0.51
Butyl acetate	8.06	2.36	-2.21
Hexanal	7.32	-1.77	-5.89
Isopentyl acetate	-4.75	1.37	-0.06
Butanol	1.48	4.73	-1.51
Butyl butyrate	-5.34	2.84	-1.11
Butyl 2-methylbutyrate	-4.65	7.66	-0.83
2-Methyl-1-butanol	3.90	-0.37	0.80
Pentanol	-5.01	1.55	3.29
Hexyl acetate	3.16	-5.33	0.63
6-Methyl-5-hepten-2-one	-0.68	-8.08	5.72
Hexanol	-1.55	-10.76	10.75
Hexyl 2-methylbutyrate	2.62	-3.99	2.26
Trans 2-hexen-1-ol	3.65	1.14	-4.48

contents of butyl butyrate ($c_{ij}=-5.34$) and 1-pentanol ($c_{ij}=-5.01$). However, the genotypes belonging to class 2 would be characterized by high contents of ethyl butyrate and butyl 2-methylbutyrate ($c_{ij}=8.60$ and 7.66 , respectively) and very low contents of 6-methyl-5-hepten-2-one and 1-hexanol ($c_{ij}=-8.08$ and -10.76 , respectively). Regarding class 3, the aromas with the highest weights in terms of their absolute values were hexanal ($c_{ij}=-5.89$), 6-methyl-5-hepten-2-one ($c_{ij}=5.72$) and 1-hexanol ($c_{ij}=10.75$). It is noticeable that whilst hexanol has a high negative weight for class 2 (-10.76), it has a high positive weight for class 3 (10.75), which could induce a great differentiation between both classes as it can be seen in the SOM map.

Soft Independent Modelling of Class Analogy

Currently, SIMCA is one of the best-known classification techniques. It computes N principal component models, one for each defined class or category, and it defines N subspaces where each object is projected. The object assignment is obtained by comparing the distances of the objects from the class models.

There were the same number of principal components ($N=14$) for each of the three classes, with those selected being those with accumulated variances higher than 95 %. A high number of factors were necessary in order to minimize type I and type II errors. A normal range for computing the SIMCA models, which were validated by cross-validation using five cancellation groups, was chosen.

Table 3 shows the classification matrix for each of the three categories. The total classification ability was 96 %. However, the total prediction ability was only 70 %. The low number of examples in some classes (such as class 3) could explain the decrease in the prediction capacity in comparison to the classification one.

The type I error, or sensitivity, was computed for each SIMCA model. Sensitivity is related to the ability

Table 3 Classification matrix

True class	Assigned class		
	1	2	3
1	39	1	0
2	2	32	0
3	1	0	15

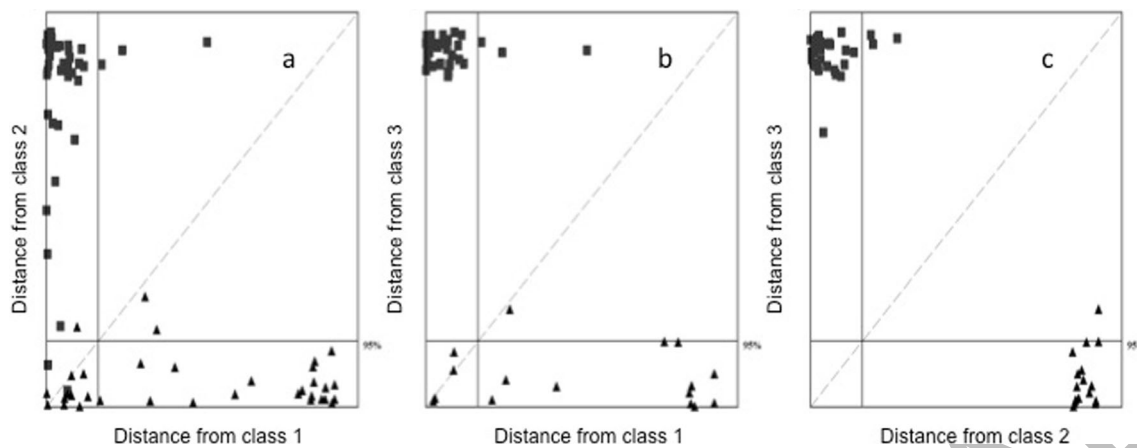


Fig. 3 Coomans' diagrams for the models constructed: **a** models 1 and 2, **b** models 1 and 3 and **c** models 2 and 3

of the model to recognize its own objects. The type II error, or specificity, was also computed for each SIMCA model. Specificity is related to the ability of the model to reject the objects that do not belong to it.

The mean sensitivity was 93.3 %, and the sensitivities computed for each class were as follows: 92.5 % for class 1, 94.1 % for class 2 and 93.8 % for class 3. On the other hand, the mean specificity was 90.6 %, and the specificities for each class were as follows: 70 % for class 1, 96.4 % for class 2 and 100 % for class 3.

Figure 3a, b, c displays Coomans diagrams for the models constructed. As seen, all of the models computed showed high sensitivities (>90 %); thus, less than 10 % of the observations belonging to each model were rejected by their own model. The class 1 model showed the worst result for the specificity parameter because it accepted 11 objects belonging to class 2 (Fig. 3a) and 4 objects belonging to class 3 (Fig. 3b). Class 2 exhibited a high specificity, as only 4 % of the cases belonging to other classes were accepted for model 2 (Fig. 3a, c). The maximum specificity was computed for class 3 (100 %) because it rejected all of the objects belonging to classes 1 and 2 (Fig. 3b, c). This fact is related to the computed distance between categories. As shown in Fig. 3 classes 1 and 2 are nearer to each other in the N-14 dimensional space than they are to class 3.

The discriminant power of each original variable (volatile compound) was also computed. The most discriminating variables were 1-pentanol, hexanal and isopentyl acetate. These results are in agreement with the PCA analysis, as these volatile compounds were correlated to genotypes belonging to classes 1 and 2, and, on the other hand, they were not correlated to genotypes belonging to class 3.

The genotypes belonging to class 3 were correlated to several alcohols such as 1-butanol, 1-hexanol and 2-methyl-1-butanol. Aliphatic alcohols are biosynthesized from fatty acids by means of a beta-oxidation mechanism, and branched alcohols are derived from amino acids, such as isoleucine in the case of 2-methyl-1-butanol. Other aromas, such as hexanal, had a low loading for the first principal component and a high loading for the third principal component (see Fig. S2 in Supporting Information). Therefore, this aroma was less related to cases belonging to breeding line 3 because these genotypes had high scores for the first principal component and low scores for the third principal component. The highest polyphenol contents of class 3 genotypes could explain this result. Hexanal is biosynthesized from linoleic acid by means of the lipoxygenase system, which is inhibited by polyphenols (Chedea and Jisaka 2011). Ethyl esters such as propanoate, 2-methylbutyrate and butyrate and acetate esters, including butyl and isopentyl, were associated to genotypes belonging to classes 1 and 2 that have low phenolic contents. Formerly, aroma types of apple genotypes were classified based on the relationship between their ester and alcohol contents (Ulrich et al. 2009).

Conclusions

This study demonstrated that the genitors used to obtain cider apple genotypes with biotic resistance, regular bearing or high polyphenol contents and late ripening have an influence on their technological and sensory properties. Thus, for example, breeding line 3 produced

genotypes richer in phenolic compounds and alcohols than genotypes obtained from breeding lines 1 and 2, whilst genotypes from lines 1 and 2 exhibited a greater quantity of acetate esters and ethyl esters than those of breeding line 3. Thus, the majority of the new genotypes obtained from breeding lines 1 and 2 could present more fruit-like odours and floral smells than those with higher polyphenol contents.

Compliance with Ethical Standards

Funding This study was funded by Instituto Nacional de Innovación Agraria (grant number INIA-12-RTA2012-00118-C03-02) and by Ministerio de Economía y Competitividad (Spain) (grant number TIN2015-65069-C2-2-R).

Conflict of Interest All authors declare that they do not have conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent Not applicable.

References

- Arias Abrodo P, Díaz Llorente D, Junco Corujedo S, Dapena de la Fuente E, Gutiérrez Álvarez MD, Blanco Gomis D (2010) Characterisation of Asturian cider apples on the basis of their aromatic profile by high-speed gas chromatography and solid-phase microextraction. *Food Chem* 121:1312–1318
- Baldwin EA (2004) Flavor. The commercial storage of fruits, vegetables, and florist and nursery stocks. Agricultural Research Service, Washington
- Ballabio D, Todeschini R (2009) Multivariate classification for qualitative analysis. Infrared spectroscopy for food quality analysis and control. 83–104. Elsevier, Amsterdam
- Blanco-Gomis D, Mangas-Alonso JJ, Junco-Corujedo S, Gutiérrez-Álvarez MD (2009) Characterization of sparkling cider by the yeast type used in taking foam on the basis of polypeptide content and foam characteristics. *Food Chem* 115:375–379
- Bus VGM, Eric van de Weg W, Peil A, Dunemann F, Zini E, Laurens FND, Blažek J, Hanke V, Forsline PL (2012) The role of Schmidt ‘Antonovka’ in apple scab resistance breeding. *Tree Genet Genomes* 8:627–642
- Chedea, V. S., Jisaka, M. (2011). Inhibition of soybean lipoxygenases. Structural and activity models for the lipoxygenase isoenzymes family. In “Recent Trends for Enhancing the Diversity and Quality of Soybean Products”, Dora Krezova (Ed), ISBN: 978-953-307-533-4
- Chen D-F, Ji Q-C, Zhao L, Zhang H-C (2009) The classification of wine based on PCA and ANN. *Advances in Intelligent and Soft Computing* 62:647–655
- Costa F, Cappellin L, Zini E, Patocchi A, Kellerhals M, Komjanc M, Gessler C, Biasioli F (2013) QTL validation and stability for volatile organic compounds (VOCs) in apple. *Plant Sci* 211:1–7
- Cozzolino D, Cynkar WU, Shah N, Damberg RG, Smith PA (2009) A brief introduction to multivariate methods in grape and wine analysis. *Int J Wine Res* 1:123–130
- Dapena E (1993) El cultivo del manzano. In: Sidra y manzana de Asturias. Ed. Prensa asturiana, Oviedo, pp 341–356
- Dapena E, Blázquez MD (2004) Improvement of the resistance to scab, rosy apple aphid and fire blight in a breeding programme of cider apple cultivars. *Acta Hort* 663:725–728
- Daszykowski M, Kaczmarek K, Vander Heyden Y, Walczak B (2007) Robust statistics in data analysis. A review basic concepts. *Chemometr Intell Lab* 85:203–219
- Díaz Llorente D, Arias Abrodo P, Dapena de la Fuente E, Mangas Alonso JJ, Gutiérrez Álvarez MD, Blanco Gomis D (2010) A novel method for the determination of total 1,3-octanediols in apple juice via 1,3-dioxanes by solid-phase microextraction and high-speed gas chromatography. *J Chromatogr A* 1217:2993–2999
- Díaz-Llorente D, Arias-Abrodo P, Dapena de la Fuente E, González-Álvarez J, Gutiérrez-Álvarez MD, Blanco Gomis D (2011) Experimental design applied to the analysis of volatile compounds in apple juice by headspace solid-phase microextraction. *J Sep Sci* 34:1293–1298
- Díaz-Llorente D, Arias-Abrodo P, González-Álvarez J, Dapena de la Fuente E, Mangas Alonso JJ, Gutiérrez Álvarez MD, Blanco Gomis D (2012) A new analytical method to volatile compounds in cider apples: application to evaluate the starch index. *Food Bioprocess Technol* 6:2447–2454
- Dixon J, Hewett EW (2000) Factors affecting apple aroma/flavour volatile concentration: a review. *N Z J Crop Hortic Sci* 28:155–173
- Dunemann F, Ulrich D, Boudichevskaia A, Grafe C, Weber WE (2009) QTL zapping of aroma compounds analysed by headspace solid-phase microextraction gas chromatography in the apple progeny ‘Discovery’ x ‘Prima’. *Mol Breeding* 23(3):501–521
- Fan W, Xu Y, Han Y (2011) Quantification of volatile compounds in Chinese ciders by stir bar sorptive extraction (SBSE) and gas chromatography mass spectrometry (GC-MS). *J Institute Brewing* 117: 61–66
- Fernández K, Labarca X, Bordeu E, Guesalaga A, Agosin E (2007) Comparative study of wine tannin classification using Fourier transform mid-infrared spectrometry and sensory analysis. *Appl Spectrosc* 61(11):1163–1167
- Fuhrmann E, Grosch W (2002) Character impact odorants of the apple cultivars Elstar and Cox Orange. *Molecular Nutr Food Res* 46:187–193
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten I (2009) The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter* 11(1):10–18
- Hibbert DH, Armstrong N (2009) An introduction to Bayesian methods for analyzing chemistry data part II: a review of applications of Bayesian methods in chemistry. *Chemometr Intell Lab* 97:211–220
- Janick J, Cummins JN, Brown SK, Hemmat M (1996) Apples. Fruit breeding: tree and tropical fruits, Vol 1. John Wiley & Sons, New York, pp 1–77
- Kafkas E, Ozogun S, Atay N, Guney M, Khodaeiaminjan M, Kafkas S (2013) Fruit quality characteristics of local and foreign apple genotypes using HPLC and HS-GC/MS techniques. *Int J Biol Pharmacy Allied Sci* 2(3):563–575
- Kohonen T (1982) Self-organized formation of topologically correct feature maps. *Biol Cybern* 43(1):59–69
- Kumar S, Volz RK, Alspach PA, Bus VGM (2010) Development of a recurrent apple breeding programme in New Zealand: a synthesis of results, and a proposed revised breeding strategy. *Euphytica* 173: 207–222
- Lea AGH, Piggot JR (1995) Fermented beverage production. Blackie Academic and Professional, London, pp 66–94
- Pawliszyn J (2012) Handbook of solid-phase microextraction. Elsevier, London., Solid-Phase Microextraction in Perspective
- Rosenblatt F (1958) The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 65(6):386–408
- Rowan DD, Hunt MB, Alspach PA, Whitworth CJ, Oraguzie NC (2009a) Heritability and genetic and phenotypic correlations of apple (*Malus domestica*) fruit volatiles in a genetically diverse breeding population. *J Agric Food Chem* 57:7944–7952

- Rowan DD, Hunt MB, Dimouro A, Alspach PA, Weskett R, Voltz RK, Gardiner S, Chagne D (2009b) Profiling fruit volatiles in the progeny of a 'Royal Gala Granny Smith' apple (*Malus domestica*) cross. *J Agric Food Chem* 57(17):7953–7961
- Sampaio K, Biasoto A, Da Silva M (2014) Comparison of techniques for the isolation of volatiles from cashew apple juice. *J Sci Food Agric* 95(2):299
- Schwab W, Davidovich-Rikanati R, Lewinsohn E (2008) Biosynthesis of plant-derived flavor compounds. *Plant J* 54:712–732
- Troggio M, Gleave A, Salvi S, Chagné D, Cestaro A, Kumar S, Crowhurst RN, Gardiner SE (2012) Apple, from genome to breeding. *Tree Genet Genomes* 8:509–529
- Ulrich D, Hoberg E, Fischer C (2009) Diversity and dynamic of sensory related traits in different *apple* cultivars. *J Applied Botany Food Quality* 83:70–75
- Wehrens R, Buydens LMC (2007) Self-and super-organizing maps in R: the Kohonen package. *J Statistical Soft* 21(5):1–19
- Zini E, Biasioli F, Gasperi F, Mott D, Aprea E, Mark TD, Patocchi A, Gessler C, Komjanc M (2005) QTL mapping of volatile compounds in ripe apples detected by proton transfer reaction-mass spectrometry. *Euphytica* 145:269–279

FOR APPROVAL