

Classifying different Iberian pig genetic lines by applying chemical–instrumental parameters of dry-cured Iberian shoulders

D. Caballero¹  · M. Asensio¹ · C. Fernández¹ · N. Martín¹ · A. Silva¹

Revised: 8 August 2018 / Accepted: 14 August 2018 / Published online: 1 September 2018
© Association of Food Scientists & Technologists (India) 2018

Abstract Iberian pigs are an autochthonous porcine breed mainly bred in southwestern of Iberian Peninsula, this herd traditionally feeds on acorn and pasture, in an extensive production system. In this study, the main objective was for classifying Iberian shoulders from different genetic lines of pigs and their crossed genetic lines. For that, morphology parameters, fatty acid profile and volatile compounds were determined on two batches of dry-cured Iberian shoulders. The samples were acquired from the same place of the dry-cured Iberian shoulders. From these data, two databases were created, for training and validation purpose. Results on this study, firstly demonstrate the capability of data mining techniques in order to classify shoulder as function as Iberian pig genetic lines by using of different quality characteristics. Four classification models were developed and tested in the training database and all models achieved a percentage of correct classification higher than 75%. Then, the same four classification models were performed in the validation database and the model of J48 decision tree and fatty acid profile reached the highest correct

classification percentage (91.67%), and from Retinto genetic line (95.65%). From this classification model, a software application was developed in order to determine in a semi-automatic way the genetic line of the dry-cured Iberian shoulder. This application could be used in the meat industries for determining relationships between genetic line and meat quality or performance of noble parts.

Keywords Retinto · Torbiscal · Decision trees · Fatty acid profile · Aromatic compounds · Morphometric measurements

Introduction

Iberian pigs are an autochthonous porcine breed mainly bred in southwestern of Iberian Peninsula, this herd traditionally feeds on acorn and pasture, in an extensive production system. The quality of meat and meat products from Iberian pigs have been amply studied for several authors (Gandermer 2009; Morcuende et al. 2003; Silva et al. 2013; Ventanas et al. 2001).

Taking into account the genetic basis, it is very important in order to obtain a genetic line that shows a greater prolificacy and performance of noble parts, without affecting the product quality. In the same breed, when comparing different genetic lines, exist variations in parameters that define the quality of fresh and dry-cured meat products derived from pigs, and, consequently these variations affect to the final quality of these products (Fernandez et al. 2012). In the case of Iberian pig, four different genetic lines (Entrepelado, Lampiño, Retinto and Torbiscal) were studied and some quality, physico-

✉ D. Caballero
dcaballero@unex.es

M. Asensio
mariaac@unex.es

C. Fernández
carlosfb@unex.es

N. Martín
noeliamn@unex.es

A. Silva
asilvaro@unex.es

¹ Animal Source Foodstuffs Innovation Services (SiPA), University of Extremadura, Av/Universidad, S/N, 10003 Cáceres, Cáceres, Spain

chemical and sensory traits were characterized (Estévez et al. 2003; Muriel et al. 2004; Tejada et al. 2002).

The influence of monounsaturated fatty acids (MUFA) on the lipid composition of subcutaneous and intramuscular fat on Iberian pigs were studied (Muriel et al. 2004; Tejada et al. 2002). Besides, the proportion of major fatty acids (FA) from subcutaneous fat were used to classify Iberian pigs as a function of their feeding background during the fattening period (Pérez-Palacios et al. 2010).

In the other hand, the generation of volatile compounds in meat and meat products were studied because of the role of flavor in the acceptability of this product (Mottram 1998). The solid phase micro extraction (SPME) combined with gas chromatography and mass spectrometry (GC–MS) were successfully used to collect aroma components in cooked pork meat and other meat products (Machielis and Istasse 2002).

Several data sets could be obtained for applying several quality parameters analyses, such as physico-chemical, FA profile and volatile compounds analyses (Silva et al. 2013). Frequently, significant information is hidden in these data sets, which could be found by using Knowledge Discovery in Databases (KDD) techniques (Fayyad et al. 1996). Data mining is an important part of KDD, mainly related to the non-trivial process of finding knowledge and potentially useful information from data stored. It is a field, which tries to describe and to predict data by exploring and analyzing large volumes of data (Sayad. 2011).

To the best of our knowledge, in the scientific literature, some studies applying classification techniques of data mining on food. Holmes et al. (2012) applied data mining techniques to detect fruits and vegetables contaminated with pesticide and identified these products as a function of their origin country. Batista et al. (2012) classified honey samples as a function of their chemical elements. Samples of confit cod were classified as a function of cooking conditions (Pérez-Palacios et al. 2017). Silva et al. (2013) classified dry-cured shoulders samples in three quality categories. Recently, Caballero et al. (2016) classified ham samples as a function of their salting stage and as a function of the feeding of the pigs (Caballero et al. 2018). However, in the scientific literature, the genetic lines did not classified as a function of morphology, FA profile and volatile compounds attributes by using data mining techniques.

This work has two main objectives. First, this study aims to classify and to characterize shoulders from two genetic lines from Iberian pigs (Retinto and Torbiscal) and their crossed genetic line as a function of their FA, volatile compounds and morphology traits by using of data mining techniques. This study aims also, developing a software application that can carried out the classification automatically with a high rate of accurate.

Materials and methods

Experimental design

This study was carried out with 90 dry-cured Iberian shoulders (30 for training batch and 60 for validation batch). For these three different Iberian pig genetic lines, 33 of them from Retinto genetic line (10 for training batch and 23 for validation batch), 30 of them from Torbiscal genetic line (10 for training batch and 20 for validation batch) and 27 from cross genetic line between Retinto and Torbiscal (10 for training batch and 17 for validation batch). These dry-cured Iberian shoulders are from the Iberian pig of the same group. These animals were castrated male Iberian pigs reared outdoors with free access to acorns during the fattening period. The slaughtering weight of these pigs were between 140 and 160 kg. The ripening period of dry-cured Iberian shoulders were around 24 months and the ripening was carried out in the same industrial dryer building (Jamon y Salud, Llerena, Spain).

For all dry-cured Iberian shoulders, firstly, morphology measures were taken. For each shoulder the length, width, perimeter and the weight were measured. Then, to always take the same portion of sample the tip of the dry-cured shoulder was taken as the reference, making the cut 9 cm from the tip and the second at 2 cm. Later, FA profile by using GC-FID and volatile compounds analysis by SPME and GC–MS were performed with samples from all piece of dry-cured Iberian shoulders.

Gathering all data from physical and instrumental analysis, two databases were created. One of them for training purpose, creating the classification model from this database and the second one for validating the same models obtained from the training batch. For evaluating the goodness of these models, the percentage of correct classification was applied.

Finally, a software application was developed based on the best classification model, allow classifying the dry-cured Iberian shoulders as a function of chemical–instrumental parameters in a semi-automatic way.

Figure 1 shows the experimental design of this study.

Fatty acid profile

Intramuscular total lipids from homogenized samples were extracted according to described by Pérez-Palacios et al. (2008). Fatty acid methyl esters (FAME) from extracted fat were prepared by acidic transesterification in the presence of sulphuric acid (5% sulphuric acid in methanol) (Sandler and Karo 1992). FAMES were analyzed by gas chromatography using a Hewlett-Packard HP-6890A gas chromatograph, equipped with a flame ionisation detector

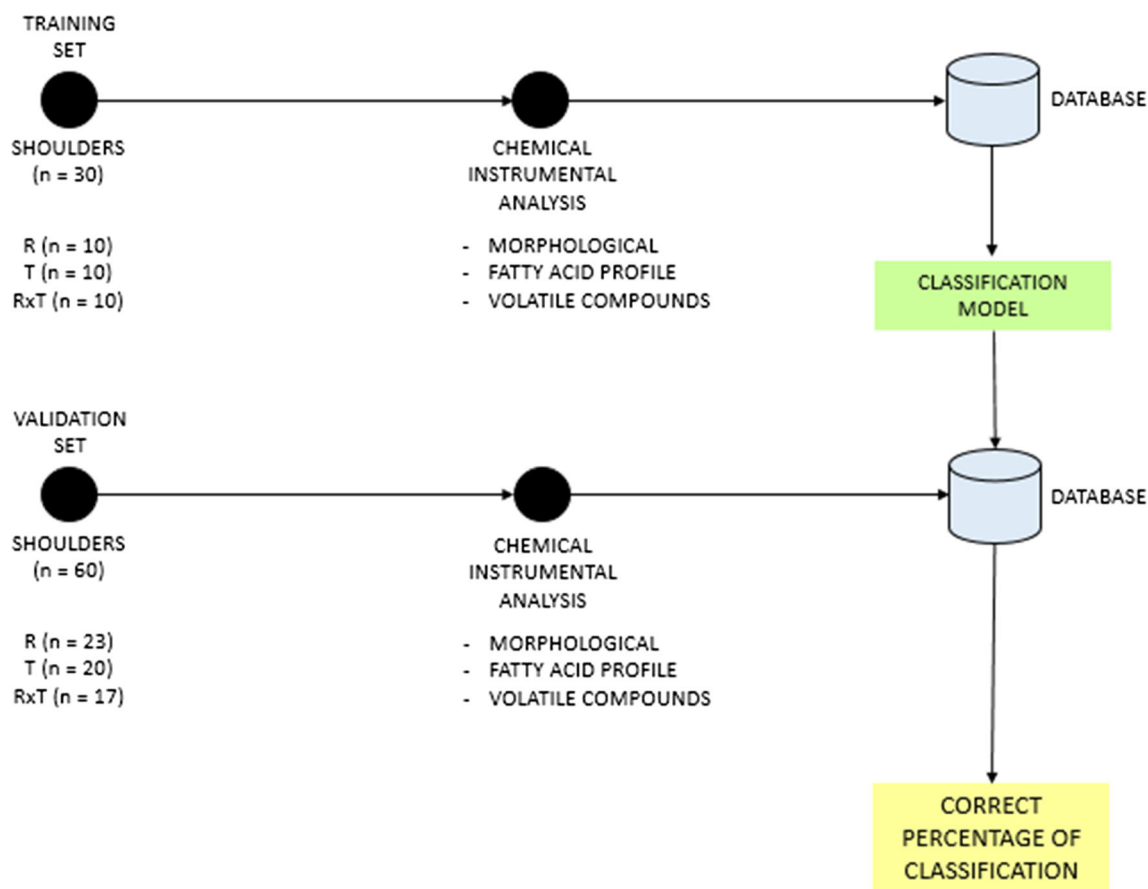


Fig. 1 Experimental design (R: Retinto, T: Torbiscal)

(FID). Separation was carried out on a polyethylene glycol-TPA modified fused silica semi-capillary column (60 m long; 0.32 mm. id, 0.25 μm film thickness) (Supelco, Bellefonte, Pennsylvania, U.S.A.). The temperature program was isothermal for 5 min at 160 $^{\circ}\text{C}$ and then raised at the rate of 5 $^{\circ}\text{C min}^{-1}$ to 250 $^{\circ}\text{C}$, and held for 5 min. Injector and detector temperatures were 250 $^{\circ}\text{C}$. The injector port was in the splitless mode. Carrier gas was nitrogen at a flow rate of 1.3 mL min^{-1} .

Individual FAMES were identified by comparing their retention time with those of standards of all fatty acid analysed supplied by Sigma Aldrich (Steintein, Germany). Peak areas were measured and percentage of FAMES were calculated basing on the relative area percentages of the FAME analysed.

Volatile compounds analysis

The SPME fibre, coated with a carboxen-poly(dimethylsiloxane) (CAR/PDMS) (70 μm thickness) was preconditioned prior analysis at 220 $^{\circ}\text{C}$ during 30 min. The headspace (HS) sampling was performed as follows: 1 g minced ham was placed in 2.5 mL vials and the SPME

fibre was exposed to the HS of the minced shoulder while the sample equilibrated during 30 min immersed in water at 37 $^{\circ}\text{C}$. Analyses were performed on a HP 6890 GC series II gas chromatograph (Hewlett-Packard, Palo Alto, California, U.S.A.) coupled to a mass-selective detector (Agilent model 5973, Agilent, Santa Clara, California, U.S.A.).

Volatile compounds were separated using a 5% phenyl-95% dimethyl polysioxane column (Restek, Bellefonte, Pennsylvania, U.S.A.) (30 m \times 0.25 mm id, 1.0 mm film thickness) (Ventanas et al. 2007). The carrier gas was Helium at 18.5 psi, resulting in a flow of 1.2 mL min^{-1} . The SPME fibre was desorbed and maintained in the injection port at 220 $^{\circ}\text{C}$ during the whole chromatography run. The injector port was in the splitless mode. The temperature program was isothermal for 10 min at 40 $^{\circ}\text{C}$ and then raised at the rate of 7 $^{\circ}\text{C min}^{-1}$ to 250 $^{\circ}\text{C}$, and held for 5 min. The GC/MS transfer line temperature was 270 $^{\circ}\text{C}$. The mass spectrometer operated in the electron impact mode with an electron energy of 70 eV, a multiplier voltage of 1650 V and collecting data at a rate of 1 scan s^{-1} over a range of m/z 40–300. The identification was carried out by comparing with NIST library and

calculation the linear retention indexes (LRI) using hydrocarbons standards. Results from the volatiles compounds analysis are provided in area units (AU).

Data mining analysis

A classification technique of data mining was applied in this study by using the free software Waikato Environment for Knowledge Analysis (WEKA) (<http://www.cs.waikato.ac.nz/ml/weka>—last accessed: August 2018).

Decision tree (DT) was used as classification technique in our study. DT is a decision modeling tool that graphically displays the classification process of a given input for given output class labels (Drazin and Montag 2012). This method is one of the learning algorithms that generate classification models in form of tree structure. It is based on the “divide and conquer” strategy (Safavian and Landgrebe 1991). Data subsets were created by decomposing the whole dataset in smaller datasets. The final model is a tree structure with decision nodes and leaf nodes, that showed the class labels for the data.

The J48 decision tree-inducing algorithm is a WEKA implementation of the well-known C4.5 decision tree. According to Anyanwu and Shiva (2009) and Priyam et al. (2013) the J48 provides better accuracy and efficiency than other decision tree algorithms. Therefore, the J48 was used as decision tree in this study. A confidence factor of 0.5 and minimum bucket size of 30 were applied (Drazin and Montag 2012).

DT was performed in each training subset for morphology measures, FA profile, volatile compounds and all together databases to create one model for each database and these models were validated on the validation subset.

The percentage of correct classification was applied to validate the goodness of the results.

Databases

A first database was created with data obtained from the results of morphological measures (length, width, perimeter and weight), percentages of FA from FA profile and volatile compounds from volatile compounds analysis from dry-cured Iberian shoulders of different training batches for all genetic lines. Besides, for each genetic line, 10 samples were used.

The second database was made with data obtained throughout the same analyses than the training batch, but the analyzed dry-cured Iberian shoulders belong to the validation batches that they have a size of 23, 20 and 17 for the Retinto, Torbiscal and their crossed genetic line, respectively. In this database, the classification models developed with the data from the training database were applied in order to validate them.

Statistical analysis

The effect of genetic lines of Iberian pig on values from morphological measures, FA profile and volatile compounds analysis were determined by one way analysis of variance (ANOVA) on training and validation datasets. When the significant differences ($p < 0.05$) were found, the Tukey's test was performed. The outliers data were carried out by using the Grubb's test. Analysis were done by using the SPSS package (v 20.0) (IBM Co., New York, New York, U.S.A.).

Results and discussion

Morphological measurements

Table 1 shows the results from the morphological measurements of dry-cured Iberian shoulders from training and validation batch.

Table 1 Average and standard deviation values of morphology characteristics of dry-cured shoulders and the effect among the different evaluated genetic lines

Genetic line	Weight	Length	Width	Perimeter	L/W ratio	L/P ratio
<i>Training batch</i>						
R	4.94 ± 0.49	67.05 ± 2.19	22.70 ± 1.11	53.36 ± 2.54	2.96 ± 0.15	1.26 ± 0.06
T	4.87 ± 0.45	68.00 ± 2.46	22.75 ± 0.86	53.50 ± 1.78	2.99 ± 0.13	1.27 ± 0.05
R × T	4.98 ± 0.45	68.00 ± 2.74	22.15 ± 1.11	53.04 ± 2.04	3.07 ± 0.12	1.28 ± 0.05
<i>p</i>	0.866	0.617	0.367	0.887	0.158	0.625
<i>Validation batch</i>						
R	4.79 ± 0.44	67.18 ± 2.68	23.36 ± 1.56	54.73 ± 2.55	2.89 ± 0.23	1.23 ± 0.08
T	4.75 ± 0.42	67.88 ± 1.86	23.30 ± 1.01	54.11 ± 1.79	2.92 ± 1.30	1.26 ± 0.05
R × T	4.87 ± 0.41	67.41 ± 2.67	22.85 ± 1.64	53.91 ± 2.60	2.96 ± 0.22	1.25 ± 0.07
<i>p</i>	0.693	0.649	0.500	0.513	0.508	0.396

R Retinto, T Torbiscal, R × T crossed line between Retinto and Torbiscal, L/W ratio ratio between length and width, L/P ratio ratio between length and perimeter

For the training batch, the average values of weight were between 4.87 and 4.98 kg. In the length case, the average values of the shoulders varied between 67.05 and 68.00 cm. For the width, the average values were found between 22.15 and 22.75 cm. The average values of the perimeter were between 53.04 and 53.50 cm. The ratio between length and width were found between 2.95 and 3.07 and for the ratio between length and perimeter varied between 1.25 and 1.28.

For the validation batch the weight were between 4.75 and 4.87 kg. For the length of the shoulders, the average values were between 67.18 and 67.88 cm, the average values for the width were found between 22.85 and 23.37 cm. The average values of the perimeter of shoulders varied between 53.90 and 54.73 cm. The ratios, between length and width, the averages values were found between 2.88 and 2.97, and between length and perimeter, the average values were between 1.23 and 1.26. In all cases, no significate differences were found ($p > 0.05$).

These results are usual among dry-cured Iberian shoulders, since all studied genetic lines come from the Iberian breed. Besides, these results are quite in concordance with previous studies (Muriel et al. 2004; Reina et al. 2012), since that indicated values around of 5 kg for the weight, between 60 and 75 cm for the length, between 20 and 25 cm for the width, between 50 and 60 cm for the perimeter. For the ratios, between length and width, the values should be between 2.40 and 3.75, in the case of the ratio between length and perimeter the values should be between 1 and 1.50.

Fatty acid profile

Table 2 shows the FA profile of the intramuscular fat for the training and validation batches of Iberian shoulders from different genetic lines.

Evaluating the results of the training batch, noting that significate differences were found ($p < 0.05$) on some FA between different genetic lines: capric acid (C10:0), myristic acid (C14:1), palmitoleic acid (C16:1), stearic acid (C18:0) and oleic acid (C18:1). Besides, significate differences were also found between genetic lines for saturated fatty acid (SFA), MUFA, polyunsaturated fatty acid (PUFA), ratio n-6/n-3 and ratio MUFA/SFA.

In the case of the results of the validation batch, noting that significate differences were found ($p < 0.05$) between different genetic lines on some FA: capric acid (C10:0), myristic acid (C14:1), palmitoleic (C16:1), stearic acid (C18:0), oleic acid (C18:1), linoleic acid (C18:2) and linolenic acid [C18:3 (n-3)]. For the groups of FA significate differences were found on: SFA, MUFA, PUFA, n-6 and n-3 group. In the case of the ratios, significate

differences were found between ratio n-6/n-3 and ratio MUFA/SFA.

In general, comparing results among different genetic lines, for the palmitic acid (C16:0) the crossed genetic line showed a concentration higher than the other two genetic lines. For the stearic acid (C18:0) the Retinto genetic line reached a concentration 2 percentage points lower than the other two genetic lines, while, in the case of the oleic acid (C18:1) the Retinto genetic line achieved a concentration between 1 and 2 percentage points higher than the others genetic lines. For the groups of FA, Retinto genetic line showed a percentage, 2 percentage points lower in the case of SFA and 2 percentage points higher in the case of MUFA than the other two studied genetic lines. Taking into account PUFA values shows levels around 9.5%, usual values from the technologic point of view.

The higher percentage of MUFA in the Retinto genetic line could invites to think about if the fat fraction is the same, in a different stimulation of the desaturated enzymes in the Iberian pigs (Muriel et al. 2004). These results could be due to a higher protein content or lower N.F.E.S. (Nitrogen-Free Extractive Substances), which, reducing the activity of the Δ^9 -desaturase, increasing the saturated and decreasing the monounsaturated fatty acid (Ventanas 2012). This fact can be verified in the ratio between oleic acid (C18:1) and stearic acid (C18:0), which is higher in Retinto genetic line, which would indicate that Δ^9 -desaturase can be more active in the Retinto than the other two genetic lines favouring the synthesis of oleic acid. These results are in concordance with other studies (Ibáñez-Escriche et al. 2016) that showed significantly values higher in Retinto genetic line than the others genetic line for MUFA group.

Volatile compounds analysis

Table 3 shows the volatile compounds for the training and validation batches of Iberian shoulders from different genetic lines.

In general, comparing the results among genetic lines for the training batch, significate differences ($p < 0.05$) were found for the 2(3H)furanone, 5-ethylidihydro compound. In the case of the validation batch, no significate differences were found for all volatiles compounds.

The results of volatile compounds are quite in concordance with previous studies about volatile components in dry-cured Iberian meat products (García et al. 1991; Ruiz et al. 1998). In both batches, training and validation batch, the majority found compounds are: Hexanal, Pentanal, Butanal 3-methyl and Butanal 2-methyl. The high quantity of Butanal 3-methyl could indicate that the branched chain aliphatic aldehydes are typical products of Strecker degradation of aminoacids for the Retinto genetic line, and,

Table 2 Average and standard deviation values of fatty acid profile of dry-cured shoulders and the effect among the different evaluated genetic lines

Genetic lines	Training batch				Validation batch			
	R	T	R × T	<i>p</i>	R	T	R × T	<i>p</i>
C10:0	0.06 ± 0.01	0.05 ± 0.01	0.04 ± 0.01	0.006	0.06 ± 0.01	0.05 ± 0.01	0.04 ± 0.01	0.009
C12:0	0.08 ± 0.01	0.07 ± 0.01	0.07 ± 0.01	0.322	0.07 ± 0.14	0.69 ± 0.11	0.71 ± 0.15	0.866
C14:0	1.58 ± 0.22	1.47 ± 0.23	1.58 ± 0.10	0.480	1.55 ± 0.22	1.50 ± 0.20	1.53 ± 0.16	0.794
C14:1	0.03 ± 0.01	0.02 ± 0.01	0.03 ± 0.01	0.029	0.03 ± 0.01	0.03 ± 0.01	0.03 ± 0.01	0.007
C15:0	0.04 ± 0.01	0.04 ± 0.01	0.04 ± 0.01	0.598	0.04 ± 0.01	0.04 ± 0.01	0.04 ± 0.01	0.987
C15:1	0.13 ± 0.02	0.14 ± 0.03	0.12 ± 0.04	0.422	0.14 ± 0.04	0.78 ± 2.87	0.14 ± 0.04	0.387
C16:0	27.07 ± 1.90	27.26 ± 2.08	28.39 ± 0.84	0.192	26.86 ± 1.73	26.99 ± 3.12	27.91 ± 1.16	0.296
C16:1	4.42 ± 0.67	3.67 ± 0.36	4.02 ± 0.37	0.008	4.32 ± 0.57	3.69 ± 0.57	3.91 ± 0.38	0.001
C17:0	0.19 ± 0.04	0.21 ± 0.04	0.19 ± 0.03	0.334	0.19 ± 0.03	0.21 ± 0.04	0.19 ± 0.03	0.146
C17:1	0.22 ± 0.04	0.22 ± 0.04	0.22 ± 0.04	0.935	0.23 ± 0.04	0.54 ± 1.39	0.22 ± 0.04	0.376
C18:0	9.83 ± 1.15	11.55 ± 0.64	11.03 ± 0.93	0.001	10.00 ± 1.02	12.24 ± 3.91	10.96 ± 0.81	0.015
C18:1	46.22 ± 2.53	45.15 ± 1.35	43.72 ± 1.06	0.014	45.84 ± 2.22	43.98 ± 4.81	44.21 ± 1.60	0.037
C18:2 (n-6)	7.77 ± 0.67	7.59 ± 0.82	8.09 ± 0.95	0.392	8.15 ± 0.93	7.38 ± 1.14	8.24 ± 0.86	0.016
C18:3 (n-6)	0.05 ± 0.01	0.05 ± 0.01	0.05 ± 0.01	0.245	0.05 ± 0.01	0.06 ± 0.03	0.05 ± 0.01	0.325
C18:3 (n-3)	0.32 ± 0.03	0.29 ± 0.03	0.32 ± 0.03	0.129	0.33 ± 0.04	0.29 ± 0.03	0.33 ± 0.01	0.001
C20:0	0.11 ± 0.01	0.13 ± 0.03	0.12 ± 0.01	0.179	0.12 ± 0.02	0.14 ± 0.07	0.12 ± 0.01	0.207
C20:1	0.66 ± 0.08	0.68 ± 0.11	0.63 ± 0.47	0.318	0.67 ± 0.10	0.66 ± 0.12	0.65 ± 0.06	0.751
C20:2 (n-6)	0.25 ± 0.03	0.24 ± 0.05	0.25 ± 0.03	0.911	0.27 ± 0.04	0.24 ± 0.05	0.26 ± 0.03	0.071
C20:3 (n-6)	0.11 ± 0.01	0.12 ± 0.03	0.11 ± 0.02	0.573	0.11 ± 0.03	0.13 ± 0.07	0.11 ± 0.02	0.418
C20:4 (n-6)	0.55 ± 0.08	0.69 ± 0.18	0.66 ± 0.15	0.082	0.64 ± 0.22	0.66 ± 0.18	0.65 ± 0.15	0.928
C20:5 (n-3)	0.03 ± 0.01	0.03 ± 0.01	0.03 ± 0.01	0.574	0.03 ± 0.01	0.03 ± 0.01	0.02 ± 0.01	0.283
C22:0	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.617	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.933
C22:1	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.867	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.263
C22:2 (n-6)	0.05 ± 0.02	0.04 ± 0.02	0.05 ± 0.01	0.433	0.04 ± 0.02	0.04 ± 0.02	0.04 ± 0.02	0.775
C22:6 (n-3)	0.07 ± 0.11	0.80 ± 0.02	0.78 ± 0.22	0.535	0.08 ± 0.02	0.08 ± 0.02	0.08 ± 0.02	0.968
C23:0	0.07 ± 0.01	0.08 ± 0.02	0.08 ± 0.02	0.119	0.08 ± 0.03	0.08 ± 0.02	0.08 ± 0.02	0.790
C24:0	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.409	0.01 ± 0.01	0.01 ± 0.01	0.01 ± 0.01	0.275
C24:1	0.04 ± 0.02	0.03 ± 0.02	0.03 ± 0.01	0.466	0.03 ± 0.01	0.03 ± 0.01	0.02 ± 0.01	0.547
Σ SFA	39.03 ± 2.85	40.89 ± 1.83	41.55 ± 1.53	0.038	38.98 ± 2.33	41.33 ± 2.78	40.97 ± 1.68	0.004
Σ MUFA	51.73 ± 2.74	49.94 ± 1.23	48.78 ± 1.09	0.005	51.29 ± 2.37	49.72 ± 1.93	49.20 ± 1.51	0.005
Σ PUFA	9.15 ± 0.76	9.06 ± 1.08	9.57 ± 1.15	0.048	9.63 ± 1.19	8.82 ± 1.36	9.72 ± 1.04	0.045
Σ n-6	8.67 ± 0.72	8.61 ± 1.01	9.10 ± 1.10	0.471	9.14 ± 1.14	8.37 ± 1.31	9.24 ± 0.99	0.047
Σ n-3	0.47 ± 0.04	0.45 ± 0.06	0.47 ± 0.05	0.538	0.49 ± 0.05	0.44 ± 0.06	0.48 ± 0.05	0.028
n-6/n-3 ratio	18.57 ± 0.69	19.39 ± 0.63	19.41 ± 1.05	0.044	18.85 ± 0.70	18.99 ± 1.65	19.38 ± 0.90	0.035
MUFA/PUFA ratio	5.70 ± 0.53	5.58 ± 0.63	5.17 ± 0.66	0.144	5.40 ± 0.68	5.79 ± 1.08	5.12 ± 0.62	0.054
MUFA/SFA ratio	1.34 ± 0.17	1.22 ± 0.08	1.18 ± 0.06	0.013	1.32 ± 0.14	1.21 ± 0.13	1.20 ± 0.08	0.003

R Retinto, T Torbiscal, R × T crossed line between Retinto and Torbiscal, SFA saturated fatty acid, MUFA monounsaturated fatty acid, PUFA polyunsaturated fatty acid, MUFA/PUFA ratio ratio between MUFA and PUFA, MUFA/SFA ratio ratio between MUFA and SFA

it could be related with a high ripening time that the shoulders from other two genetic lines. The importance of these compounds to the overall flavour was described previously in some products of pork meat (Hinrichsen and Andersen 1994), the high percentage found in dry-cured Iberian ham, and shoulders could be one of the reasons for

their high acceptability (Ruiz et al. 1998). Besides, straight chain aliphatic aldehydes are typical products of lipid oxidation. The high quantity of Hexanal could be related with a carbonyl that arises from the oxidation of n-6 FA group (Table 2). These results are similar to the results obtained by García et al. (1991) and Ruiz et al. (1998).

Table 3 Average and standard deviation values of volatile compounds profile of dry-cured shoulders and the effect among the different evaluated genetic lines

Genetic lines	Training batch				Validation batch			
	R	T	R × T	p	R	T	R × T	p
	2,3-butanedione	0.10 ± 0.06	0.13 ± 0.09	0.14 ± 0.06	0.372	0.15 ± 0.10	0.11 ± 0.11	6.26 ± 1.66
Butanal, 3-methyl	27.92 ± 18.92	25.49 ± 14.96	20.02 ± 7.18	0.470	25.61 ± 13.68	24.45 ± 11.09	18.69 ± 6.97	0.144
Butanal, 2-methyl	9.01 ± 3.72	9.56 ± 5.25	7.57 ± 2.47	0.520	8.49 ± 3.40	7.67 ± 4.41	8.19 ± 6.22	0.849
Pentanal	12.45 ± 5.89	20.86 ± 10.16	15.56 ± 7.26	0.347	20.25 ± 17.00	21.66 ± 16.15	30.35 ± 23.54	0.668
Hexanal	104.18 ± 70.39	138.23 ± 104.13	112.85 ± 42.21	0.592	150.83 ± 114.99	140.40 ± 85.28	108.56 ± 54.30	0.344
2-heptanone	8.20 ± 3.53	9.82 ± 5.33	6.95 ± 4.09	0.355	11.48 ± 6.45	15.35 ± 12.77	17.80 ± 15.12	0.548
Heptanal	4.84 ± 2.73	7.64 ± 6.05	5.76 ± 2.33	0.308	10.34 ± 9.58	7.82 ± 5.70	6.66 ± 4.99	0.269
2(3H)furanone, dihydro-5-methyl	0.10 ± 0.07	0.10 ± 0.08	0.09 ± 0.07	0.963	0.14 ± 0.11	0.12 ± 0.10	0.08 ± 0.06	0.322
Furan,2-pentyl	0.37 ± 0.26	1.16 ± 0.53	0.27 ± 0.13	0.656	0.56 ± 0.50	0.87 ± 0.45	0.53 ± 0.52	0.862
Octanal	2.32 ± 2.21	3.60 ± 2.40	2.45 ± 1.57	0.464	6.79 ± 6.23	3.42 ± 3.19	3.96 ± 2.53	0.155
2(3H)furanone, 5-ethylidihydro	0.04 ± 0.01	0.61 ± 0.41	0.07 ± 0.01	0.039	0.77 ± 0.39	0.47 ± 0.28	0.21 ± 0.12	0.320
Nonanal	3.20 ± 2.55	4.78 ± 2.62	5.15 ± 2.57	0.219	7.00 ± 5.42	4.80 ± 3.37	6.38 ± 3.30	0.315
Decanal	0.14 ± 0.05	0.14 ± 0.08	0.14 ± 0.11	0.624	0.14 ± 0.06	0.14 ± 0.06	0.14 ± 0.06	0.176
2(3H)furanone, 5-butylidihydro	0.06 ± 0.04	0.25 ± 0.12	0.08 ± 0.04	0.347	0.14 ± 0.06	0.14 ± 0.06	0.14 ± 0.06	0.441

R Retinto, T Torbiscal, R × T crossed line between Retinto and Torbiscal

Classification model development

Recent studies showed the goodness in order to classify of the lineal discriminant algorithms (LDA) based on trees structures such as decision trees (DT) or Random Forest (RF) (Caballero et al. 2016; Cernadas et al. 2017; Fernandez-Delgado et al. 2014). In comparison with classical techniques of classification such as K-NN, Rules Based Systems (RBS), Artificial Neural Networks (ANN) or deep learning methods, the LDA showed better results (Fernandez-Delgado et al. 2014). From all these classification techniques based on trees structure, DT was selected as classification techniques, since, DT is one of the simpler tree structure and concretely, J48 DT is one of the most efficient techniques (Anyanwu and Shiva 2009; Priyam et al. 2013).

Table 4 shows the percentage of correct classification for the training and validation batches of Iberian shoulders from different genetic lines as a function of the chemical–instrumental characteristics of them.

In general, for the training batch, the higher percentage of correct classification as a function of the genetic lines were achieved for the Torbiscal genetic line (100%), followed by the crossed line between Torbiscal and Retinto (90–100%), and, reached the lower percentage of correct classification for Retinto genetic line (80–100%). Taking into account the chemical–instrumental characteristics applied for the classification purpose, the highest percentages of correct classification were reached by using of FA profile (96.67%), volatile compounds profile (96.67%) and all characteristics joined (96.67%), obtaining the lower percentage of correct classification for morphology characteristics (93.33%). This fact could be related with the information for the morphology characteristics is lower than the other chemical–instrumental characteristics groups, only 6 characteristics for morphology and 36 and 14 respectively for the other groups of chemical–instrumental characteristics.

Since all classification models for the training batch achieved a percentage of correct classification higher than 75%, which is the minimum percentage of correct classification (Hearty and Gibney 2008). All classification models were tested with the validation batch, in order to corroborate and to validate the goodness of the generated models.

Table 4 shows the results for the validation batch, taking into account the percentages of correct classification among genetic lines of Iberian pigs, since the model was developed on data from the training batch; the percentages of correct classification for this batch are lower than the training batch. The highest percentages were obtained for the Torbiscal genetic line (90%) in general. Although, in the case of classifying the shoulders as a function of FA profile or all characteristics joined, the highest percentage of correct classification were reached for the Retinto genetic line (95.65%). Comparing the percentages of correct classification as a function of the chemical–instrumental characteristics applied to the classification purpose by using all data from the studied genetic lines, the highest correlation coefficient were reached classifying as a function of FA profile (91.67%), noting that for all characteristics, the correct classification percentage was 90%, slightly lower than applying FA characteristics. Although in all cases, the percentage of correct classification were higher than 75% (Hearty and Gibney 2008). Besides, the FA profile shows the highest number of characteristics (36) for all group of characteristics applied (6 and 14, respectively), and slightly lower than applying all characteristics joined (56). In addition, as previously was mentioned, FA profile showed a higher number of variables with significant statistically difference (Table 2) than the volatile compounds and morphology measurements group (Tables 1, 3). This fact could be related with the good results obtained for the FA profile group.

These results indicate that the application of J48 decision tree on FA profile characteristics could be used for

Table 4 Percentage of correct classification of dry-cured Iberian shoulder as a function of (a) their morphology characteristics, (b) their fatty acid profile, (c) their volatile compounds and (d) all attributes together

	Training batch				Validation batch			
	ALL (%)	R (%)	T (%)	R × T (%)	ALL (%)	R (%)	T (%)	R × T (%)
Morphology	93.33	80	100	100	81.67	73.91	90	82.35
Fatty acid profile	96.67	100	100	90	91.67	95.65	90	88.24
Volatile compounds profile	96.67	90	100	100	88.33	86.96	90	88.24
Morphology + fatty acid profile + volatile compounds profile	96.67	100	100	90	90.00	91.30	90	88.24

R Retinto, T Torbiscal, R × T crossed line between Retinto and Torbiscal

ALL: R + T + R × T

classifying the shoulders as a function of their genetic lines with a high degree of accurate.

Software application

Once the J48 DT classification models by using of morphology measures, FA profile characteristics, volatile compounds and all traits joined were tested. In order to develop a software application, it has been decided using FA profile characteristics since this model offered the highest percentage of correct classification, 96.67% for training batch and 91.67% for validation batch, and a number of characteristics (36) lower than all characteristics joined (56). Besides, the simplicity and efficiency of classification models based on trees was tested previously (Caballero et al. 2016; Cernadas et al. 2017; Fernandez-Delgado et al. 2014). Therefore, the selected model based on FA profile characteristics seems to be the best option to be implemented in a software application.

This software application was developed by using the programming language Visual C with the package Open

CV under tool Microsoft Visual Studio 2012 (Seattle, Washington, U.S.A.).

Initially, the software application shows in the screen a list with the FA and a white square for adding the percentage of this FA, this list must be completed for the user with the value of the percentage for each FA. Once the list was completed, the sum of FA groups and the ratios between groups were calculated automatically. Then, the classification model was applied automatically on this data, our model calculated the genetic line, selecting the correct branch until reaching the leaf of our model, part of this decision tree is shown in the Fig. 2b. Once our model reached a leaf of our tree, the result was shown automatically in the screen of application (Fig. 2a).

The main advantage of this classification model is that by means of this application, the meat industries could verify the genetic line of the shoulder. This application based on FA profile could discriminate the shoulders as function of their genetic line. In addition, this could be interesting for the meat industries, since the genetic line influences in some quality parameters of meat and meat products and other parameters related with the prolificacy

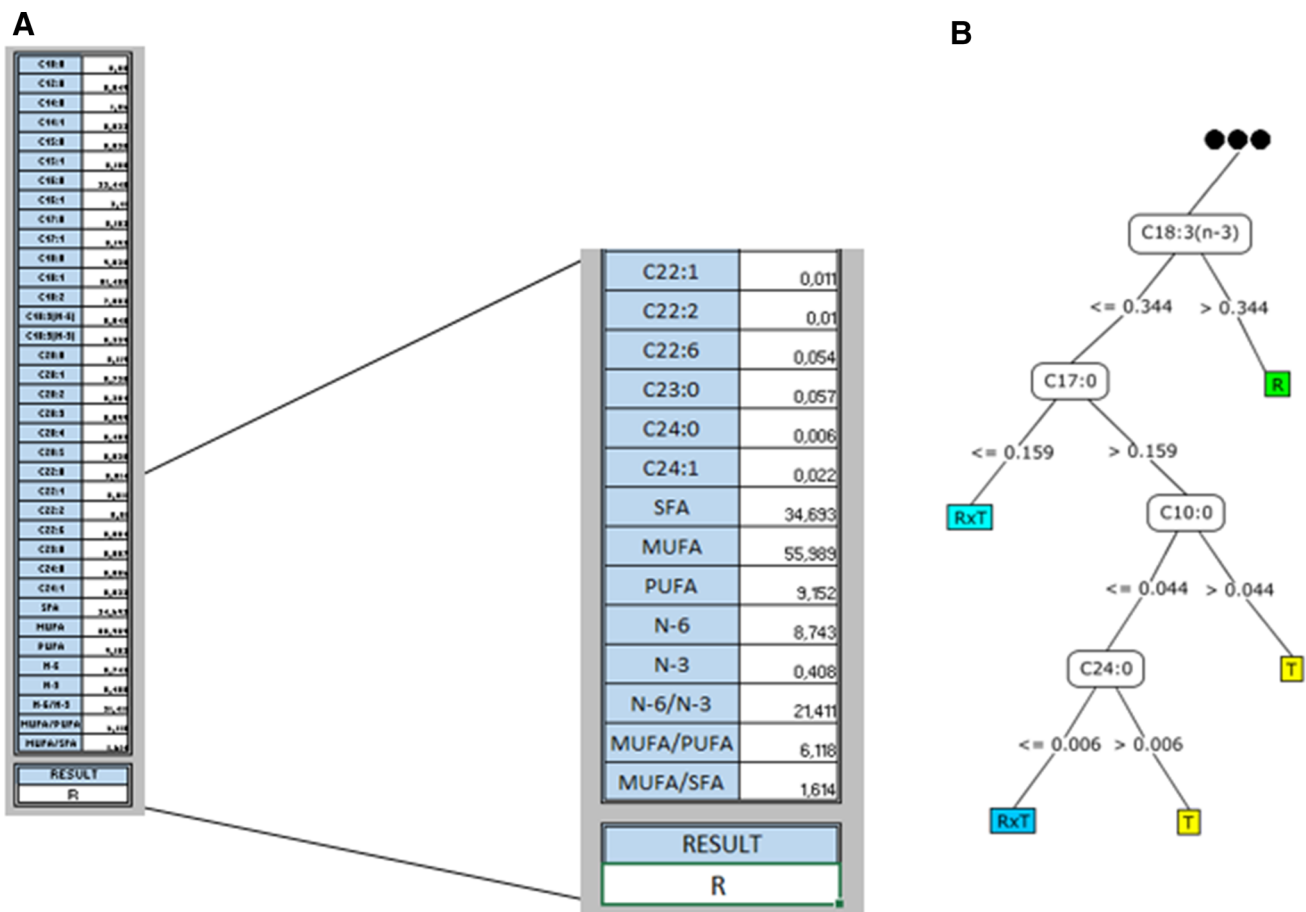


Fig. 2 a Example of the result screen of the software application. b Fragment of the applied decision tree as classification model in the software application (R: Retinto, T: Torbiscal)

and performance of the noble parts (Fernandez et al. 2012). Although, further investigations with more genetic lines such as “Entrepelado” or “Lampião” should be carried out, and with pigs with different diets in order to establish this software for the meat industries. In this way, some studies have classified samples of meat products as a function of feeding of the pigs (Caballero et al. 2018; Pérez-Palacios et al. 2010).

Conclusion

In this study, taking into account the preliminary characterization of these shoulders, from different Iberian genetic lines and their crossed lines. There are some quality indicators that influence the organoleptic evaluation of these products and, therefore, in its final quality. The Retinto genetic line stands out, which shows the best results at the level of the fatty acid profile, as the ratio between oleic acid (C18:1) and stearic acid (C18:0) or the ratio between MUFA and SFA, probably because the activity of the Δ^9 -desaturase is more intense in this genetic line.

Besides, in this study, four new models of classification were performed, showing the capability of data mining techniques and chemical–instrumental data for classifying dry-cured Iberian shoulders as a function of its genetic line. The FA profile data achieved the highest percentage of correct classification. A software application was developed based on the best model of classification tested on the validation batch in this study. This software application can discriminate samples from different genetic lines of Iberian pig in automatic way with a high rate of accurate. Currently, this application is offered to the industries by an external service of the University of Extremadura that can verify the genetic line of their dry-cured Iberian shoulders.

Acknowledgements This study was partially funded by “Centro para el Desarrollo Tecnológico Industrial” (CDTI: IDI-20100447 and CDTI: IDI-20140447). The authors also wish to acknowledge to Emilio Magallón, Manuel Ramos, Pilar Diaz, María José García and Luis Muñoz from INGAFOOD S.A. corporation (Almendralejo, Spain) and Noelia Ibáñez and José Luis Noguera from IRTA (Lleida, Spain) for his direct contribution and support.

Compliance with ethical standards

Conflict of interest The authors confirm that this article content has no conflict of interest.

References

Anyanwu MN, Shiva SG (2009) Comparative analysis of serial decision tree classification algorithms. *Int. J Comput Sci Secur (IJCSS)* 3:230–240

- Batista BL, Da Silva LRS, Rocha BA, Rodrigues JL, Beretta-Silva AA, Bonates TO, Gomes VSD, Barbosa RM, Barbosa F (2012) Multi-element determination in Brazilian honey samples by inductively coupled mass spectrometry and estimation of geographic origin with data mining techniques. *Food Res Int* 29:209–215
- Caballero D, Caro A, Rodríguez PG, Durán ML, Ávila MM, Palacios R, Antequera T, Pérez-Palacios T (2016) Modeling salt diffusion in Iberian ham by applying MRI and data mining. *J Food Eng* 189:115–122
- Caballero D, Pérez-Palacios T, Ávila MM, Caro A, Antequera T (2018) Clasificación de jamones Ibéricos mediante análisis de Imágenes de Resonancia Magnética (MRI). In: X Foro ANVEPI, Badajoz, Spain
- Cernadas E, Fernández-Delgado M, González-Rufino E, Carrión P (2017) Influence of normalization and color space to color texture classification. *Pattern Recognit* 61:120–138
- Drazin S, Montag M (2012) Decision tree analysis using WEKA. Machine learning-project II. University of Miami, Miami
- Estévez M, Morcuende D, Cava R (2003) Physico-chemical characteristics of *M. Longissimus dorsi* from three lines of free range reared Iberian pigs slaughtered at 90 kg live-weight and commercial pigs: a comparative study. *Meat Sci* 65:1139–1146
- Fayyad U, Piatetsky-Shapiro G, Smyth P (1996) From data mining to knowledge discovery in databases. *AI Mag* 17:37–54
- Fernandez J, Clemente I, Amador C, Membrillo A, Azor P, Molina A (2012) Use of different sources of information for the recovery and genetic management of endangered populations: example with the extreme case of Iberian pig Dorado strain. *Livest Sci* 149:282–288
- Fernandez-Delgado M, Cernadas E, Barro S, Amorim D (2014) Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res* 15:3133–3181
- Gandermer G (2009) Dry-cured ham quality as related to lipid quality of raw material and lipid changes during processing: a review. *Grasas Aceites* 60(3):297–307
- García C, Berdagué JJ, Antequera T, López-Bote C, Córdoba JJ, Ventanas J (1991) Volatile components of dry-cured Iberian ham. *Food Chem* 41:23–32
- Hearty AP, Gibney MJ (2008) Analysis of meal patterns with use of supervised data mining techniques – artificial neural networks and decision trees. *Am J Clin Nutr* 88:1632–1642
- Hinrichsen IL, Andersen HJ (1994) Volatile compounds and chemical changes in cured pork: role of three halotolerant bacteria. *J Agric Food Chem* 42:1537–1542
- Holmes G, Fletcher D, Reutermann P (2012) An application of data mining to fruit and vegetable sample identification using gas chromatography–mass spectrometry. In: X international congress on environmental modeling and software managing resources of a limited planet, Leipzig, Germany
- Ibáñez-Escriche N, Magallón E, González E, Tejada JF, Noguera JL (2016) Genetic parameters and crossbreeding effects of fat deposition and fatty acid profiles in Iberian pig lines. *J Anim Sci* 94:28–37
- Machiels D, Istasse L (2002) Solid-phase microextraction evaluation of key aroma compounds in cooked beef meat as influenced by breed and diet. In: 48th international congress on meat science and technology (ICOMST), Rome, Italy
- Morcuende D, Estévez M, Ruiz J, Cava R (2003) Oxidative and lipolytic deterioration of different muscle from free-range reared Iberian pigs under refrigerated storage. *Meat Sci* 65:1157–1164
- Mottram DS (1998) Flavour formation in meat and meat products: a review. *Food Chem* 62:415–424
- Muriel E, Ruiz J, Ventanas J, Petró MJ, Antequera T (2004) Meat quality characteristics in different lines of Iberian pigs. *Meat Sci* 67:299–307

- Pérez-Palacios T, Ruiz J, Tejeda JF, Antequera T (2008) Subcutaneous and intramuscular lipid traits as tools for classifying Iberian pigs as a function of their breeding background. *Meat Sci* 81:632–640
- Pérez-Palacios T, Antequera T, Durán ML, Caro A, Rodríguez PG, Ruiz J (2010) MRI-based analysis, lipid composition and sensory traits for studying Iberian dry-cured hams from pigs fed with different diets. *Food Res Int* 43:248–254
- Pérez-Palacios T, Caballero D, Bravo S, Mir-Bel J, Antequera T (2017) Effect of cooking conditions on quality characteristics of confit cod: prediction by MRI. *Int J Food Eng* 13(8):20160311
- Priyam A, Abhijeeta GR, Rathee A, Srivastava S (2013) Comparative analysis of decision tree classification algorithms. *Int J Curr Eng Tech (IJCET)* 3(2):334–337
- Reina R, López-Buesa P, Sánchez del Pulgar J, Ventanas J, García C (2012) Effect of IGF-II (insulin-like growth factor-II) genotype on the quality of dry-cured hams and shoulders. *Meat Sci* 92:562–568
- Ruiz J, Ventanas J, Cava R, Andrés A, García C (1998) Volatile compounds of dry-cured Iberian ham as affected by the length of the curing process. *Meat Sci* 52:19–27
- Safavian R, Landgrebe D (1991) A survey of decision tree classifier methodology. *IEEE Trans Syst Man Cybern* 21(3):660–674
- Sandler SR, Karo W (1992) Source book of advanced organic laboratory preparations. Academic, San Diego
- Sayad S (2011) Real time data mining. Self-Help Publishers, Cambridge
- Silva A, Reina R, García-Casco J, Ventanas J (2013) Chemical–instrumental–sensory parameters and chemometrics as tools to discriminate among the quality categories of dry-cured Iberian shoulder. *Grasas Aceites* 64(2):201–209
- Tejeda JF, Gandermer G, Antequera T, Viau M, Garcia C (2002) Lipid traits of muscles as related to genotype and fattening diet in Iberian pigs: total intramuscular lipids and triacylglycerols. *Meat Sci* 60:357–363
- Ventanas J (2012) Jamón Ibérico y Serrano: Fundamentos de la elaboración y la calidad. Mundi-Prensa, Madrid
- Ventanas J, Tejeda JF, Petró MJ (2001) El jamón Ibérico: de una imagen de calidad a una imagen definida y contrastada. In: Ventanas J (ed) *Tecnología del jamón Ibérico*. Mundi-Prensa, Madrid, pp 15–45
- Ventanas S, Ventanas J, Tovar J, García C, Estévez M (2007) Extensive feeding versus oleic acid and tocopherol enriched mixed diets for the production of Iberian dry-cured hams: effect on chemical composition, oxidative status and sensory traits. *Meat Sci* 77:246–256