

# Classification of Lentils, Chickpeas and Beans Based on Their Isoflavone Content

Lara Pérez-Martín<sup>1</sup> · Myriam Bustamante-Rangel<sup>1</sup> · M. Milagros Delgado-Zamarreño<sup>1</sup>

Received: 17 May 2016 / Accepted: 25 September 2016 / Published online: 11 October 2016  
© Springer Science+Business Media New York 2016

**Abstract** Food legumes (*Fabaceae*) form an important part of the human diet; besides, several *Fabaceae* species are acknowledged for their high levels of bioactive compounds, among which are isoflavones, being recognised for their varied types of biological activity. The aim of this work was to classify different varieties of three types of legumes (chickpeas, lentils and beans) according to their isoflavone contents. The analysis of isoflavones was carried out using high-performance liquid chromatography coupled to triple quadrupole tandem mass spectrometry (HPLC-MS/MS). To extract the analytes, a modified QuEChERS approach was used. The chromatographic peak areas obtained, after scaling in Pareto mode, were used to build statistical models. Both supervised and unsupervised techniques were applied for the classification of the different types of pulses analysed in the study: principal component analysis (PCA), hierarchical cluster analysis (HCA) and partial least squares discriminant analysis (PLS-DA). The statistical models were validated by internal validation, obtaining satisfactory results for the different matrices. PCA models allowed the differentiation between subspecies, but not subspecies, varieties or ecotypes. The results provided by HCA and PLS-DA revealed that the different species and subspecies of beans and the different varieties and subvarieties of lentils can be distinguished, and even the different ecotypes of the same variety in the case of chickpeas. This study revealed that it was possible to differentiate among species, subspecies, varieties and even ecotypes of different types of legumes based on their isoflavone content.

**Keywords** Isoflavones · Legumes · Principal component analysis · Hierarchical cluster analysis · Partial least squares discriminant analysis

## Introduction

Legumes are the second most important crop as a source of food, feed for livestock and raw materials for industry. Many of the seeds of this family (*Fabaceae*) are edible. There are 13,000 species of known legumes, of which about 20 are consumed by humans (Mazur et al. 1998). The importance of legumes is also associated with their secondary metabolites, in particular phytoestrogens. The three main classes of phytoestrogens are isoflavones, lignans and coumestans (Jacobs et al. 2009), and their biosynthesis depends strongly on the environment and plant species in question (Martin et al. 2006).

Isoflavones in particular are a subclass of flavonoids with a chemical structure similar to that of oestradiol. They are mainly found as glycosides, conjugated with carbohydrates, which may be esterified with acetyl or malonyl groups. Flavonoids and isoflavonoids have phenolic groups at both ends of the molecule, making them unique from the chemical, biochemical and biological points of view and affording the molecules exceptional stability (Adlercreutz 1998).

Depending on the methylation or glycosylation of the hydroxyl groups, the ability of isoflavones to bind to oestrogen receptors varies. Recently, the possibility of a duality in the estrogenic activity of isoflavones, which depends on the concentrations of endogenous estrogens, has been studied (Frederick and McCarty 2006; Wood et al. 2006). This underscores the beneficial effects gained from the consumption of this type of phytoestrogens for humans, such as the treatment of menopausal symptoms (Cassidy et al. 2006; Hwang et al. 2006), as an alternative to hormone-replacement therapy

✉ M. Milagros Delgado-Zamarreño  
mdz@usal.es

<sup>1</sup> Departamento de Química Analítica, Nutrición y Bromatología, University of Salamanca, 37008 Salamanca, Spain

(HRT); in cardiovascular disease (Trock et al. 2006; Rimbach et al. 2008); in diabetes and obesity (Cano et al. 2010); in osteoporosis (Morris et al. 2006); and even in cancer (e.g. prostate, Raschke et al. 2006; and endometrial cancer, Hale et al. 2002).

Isoflavones are typically associated with soybean, despite being synthesized by many other *Fabaceae* plants, such as alfalfa, clover, pea, peanut, and beans (Jacobs et al. 2009). In plants, isoflavones are found mainly as biologically inactive glycoconjugates and are hydrolysed in the intestine to their biologically active form, aglycone. The most abundant type of isoflavone depends on the legume in question. Thus, for example, soybean contains mainly daidzein, glycitein and genistein as aglycones and their glycoside, acetylglycoside and malonylglycoside forms. However, chickpeas contain biochanin-A as the major component. The isoflavone profile also is highly influenced by genotype (Lee et al. 2010; Kumar et al. 2010), agricultural practices (Vamerali et al. 2012), climatic factors (Morrison et al. 2010), processing conditions, etc. (Kumar et al. 2010).

In this work, the isoflavone profiles of different types of chickpeas, lentils and beans were obtained and evaluated using different multivariate analysis techniques in order to check how the isoflavone content is affected by the species, subspecies, variety and even ecotype of the considered legume.

This study is important for the quality authentication of legumes, as it can allow for both the classification of the different kinds of legumes and the detection of fraud by adulteration.

To carry out this study, a modified QuEChERS approach, described previously (Delgado-Zamarreño et al. 2012), was implemented for the analysis of isoflavones in legume samples, followed by separation and quantification using HPLC-MS/MS. The data from HPLC-MS/MS were three-dimensional (retention time, mass and abundance) and therefore had to be pretreated. The chromatographic peak areas obtained after scaling in Pareto mode were used to build statistical models. Both supervised (principal component analysis and hierarchical cluster analysis) and unsupervised (partial least squares discriminant analysis) techniques were applied for the different types of pulses.

## Material and Methods

### Chemicals

The following are the isoflavone standards: daidzin (CASRN552-66-9), glycitin (CAS RN40246-10-4), genistin (CASRN529-59-9), daidzein (CAS RN 486-66-8), glycitein (CASRN4095783-3), genistein (CASRN 446-72-0), formononetin (CASRN485-72-3) and biochanin-A (CAS

RN491-80-5) were purchased from Sigma-Aldrich (Alcobendas, Madrid, Spain). The organic solvents—acetonitrile (ACN), methanol (MeOH) and ethanol (EtOH)—were of HPLC grade and were supplied by Merck (Darmstadt, Germany). Formic acid (>98 %) was from Fluka (Alcobendas, Madrid, Spain). Anhydrous magnesium sulphate, sodium chloride and trisodium citrate dihydrate ( $\text{Na}_3\text{Cit}\cdot 2\text{H}_2\text{O}$ ) were from Scharlau (Barcelona, Spain). Disodium hydrogencitrate sesquihydrate ( $\text{Na}_2\text{HCit}\cdot 1.5\text{H}_2\text{O}$ ) was from Sigma-Aldrich. Ultra-high-quality (UHQ) water was obtained with a Wasserlab (Spain) water purification system. All other chemicals used were of analytical reagent grade.

### Samples

The samples analysed were purchased from local supermarkets.

Spanish chickpeas for human consumption are, for the most part, ecotypes from the macrocarpum variety of the *Cicer arietinum* L. species. Five ecotypes are accepted as such, of which only “Castellano” (CS), “Pedrosillano” (PD) and “Blanco Lechoso” (BL) were studied in this work. The chickpeas analysed were from different origins, manufacturers and harvests, resulting in 20 samples from nine different producers from the regions of Andalucía and Castilla y León (Spain). The samples analysed were labelled with the name of the producer as “Fuentesauco” (Fm, Fch, Ftg), “Pico Pardal” (Ppa, Ppv), Pedrosillano (Psc, Pbo, Pa, Pm, Prg), Blanco Lechoso (Lc, La, Lci), “Maragato” (Ma), “Gordo” (Gs), Castellano (Cv, Cci), “organically grown” (Ec) and “Mexican” (Mxa, Mxh). It should be noted that samples labelled as Fuentesauco, Pico Pardal, Maragato and Gordo, all from the region of Castilla y León, could be considered as belonging to the Pedrosillano ecotype.

The lentils most widely cultivated in Spain belong to the species *Lens culinaris* Medicus. There are three different varieties of this species: *L. culinaris* Medicus var. *variabilis*, *L. culinaris* Medicus var. *dupuyensis* and *L. culinaris* Medicus var. *vulgaris*, latter two having subvarieties. These two subvarieties are commonly commercialized as “Rubia de la Armuña” (AR) and “Rubia Castellana” (CA); the variety *variabilis* is commercialized as “Pardina” (PA) and the variety *dupuyensis* as “Verdina” (VE). Twenty lentil samples from eight different producers of these three varieties from Castilla y León (Spain) were studied. The samples analysed were labelled with the name of producer as follows: Rubia de La Armuña (Al, Ald, Ach, Li), Rubia Castellana (Ca, Cv, Cci), Pardina (Ps, Pl, Pv, Pci, Pba, Pa, Ph), Verdina (Vv), organically grown (Ec), “Extrafina” (Ea) and “Rápida” (Rh), and other lentils labelled as (Lh) and (Lp). As can be seen, some samples were of unknown variety.

Finally, two different species of beans were analysed, *Phaseolus multiflorus* and *Phaseolus vulgaris*, including two

different subspecies from the latter, *nanus* and *volubilis* spp. In this case, 20 bean samples from 12 different producers in Castilla y León and Asturias (Spain) were studied. These samples were labelled by the producer as “Blanca Riñón” (Rar, Rvg, Rd, Rla, Rls, Rl, Rvb, Rrg, Ra, Rh, Jb), “Redonda Manteca” (Rv), “Faba” (Fp, Fci, Fa, Fv), “Granja” (Gh) “Plancheta” (Pv), “Negrita” (NI) and “Judión” (Jv). These beans belong to the following varieties: Blanca Riñón (RI) and Redonda Manteca (RE) from the subspecies *nanus*; “Granja Fabada” or Faba (FA), Plancheta (PL) and Negrita (NE) from the subspecies *volubilis*, all from the species *P. vulgaris*. The species *P. multiflorus* is commercialized in Spain as Judión (J).

### Sample Preparation

Before extraction, samples were ground with a Knifetec™ 1905 from Foss (Barcelona, Spain) and were stored in a cool, dry place in closed containers.

Extraction was carried out following the QuEChERS methodology optimized in a previous work (Bustamante-Rangel et al. 2014). Depending on the type of matrix, an appropriate amount of sample was extracted in two steps. First, 10 mL of ACN/H<sub>2</sub>O (70:30, v/v) was added, and the mixture was shaken for 5 min; then, 5 mL of ACN was added and the mixture was shaken again for 5 min. The chickpea and bean extractions were carried out using a VWR thermostatted shaking tray (Radnor, PA, USA), while lentil extractions were accomplished using a Microson Misonix Ultrasonic Cell Disruptor probe (Farmingdale, NY, USA). These agitation systems were chosen taking into account the results obtained in a comparative study aimed at placing the sample and the solvent in contact in order to optimize the extraction of isoflavones from legumes using the QuEChERS method (Bustamante-Rangel et al. 2014). After extraction, a mixture of MgSO<sub>4</sub>/NaCl (4:1, w/w) together with a citrate buffer (in the case of chickpeas and beans) was added, and the mixture was shaken vigorously for 1 min and centrifuged at 3000 rpm for 5 min. The upper layer was filtered through a 0.22- $\mu$ m PVDF syringe filter before injection into the chromatographic system.

### LC-MS/MS Conditions

LC analyses were performed on a HP 1200 Series chromatograph from Agilent (Waldbronn, Germany) equipped with a binary pump, a membrane degasser, an autosampler and a six-port valve. The analytical column was a 50  $\times$  4.6-mm Zorbax Eclipse XDB-C18 with 1.8  $\mu$ m particles (Agilent). The mobile phase consisted of an acetonitrile (solvent A) and 0.01 % aqueous formic acid (solvent B) gradient at a flow rate of 0.5 mL min<sup>-1</sup>. Gradient elution was as follows: 0–1.5 min, 10 % A; 1.5–2.5 min, 10–25 % A; 2.5–3.5 min, 25 % A; 3.5–7 min, 25–50 % A; 7–8 min, 50–80 % A; 8–10 min, 80 % A;

and 10–12 min, 80–10 % A. The analytical column was thermostatted at 25 °C, and the injection volume was 2  $\mu$ L. Detection was carried out on a Triple Quad LC/MS 6410 device (Agilent) equipped with an electrospray (ESI) source. ESI-MS spectra were acquired in positive-ion multiple reaction monitoring (MRM) mode. The conditions of the MS analysis were as follows: the electrospray capillary voltage was 3500 V, and the nebulizer pressure was 35 psi. Nitrogen was used as the drying gas at a flow rate of 12 L min<sup>-1</sup> and a temperature of 350 °C. The whole system was controlled by the Agilent MassHunter software, version B.04.01.

Quantification of the isoflavones was performed using the external standard method.

### Multivariate Data Analysis

Data analysis was performed with multivariate statistical methods, using the SIMCA P+ version 13.0 software (Umetrics, Sweden). Pareto scaling was applied to the data prior to the development of the chemometric models. Principal component analysis (PCA), hierarchical cluster analysis (HCA) and partial least squares discriminant analysis (PLS-DA) were used for the discrimination and classification of legumes.

PCA is one of the most widely used unsupervised techniques in multivariate data analysis because it is easy to interpret. PCA helps to reduce the dimensionality of the data while retaining most of the information because all the relationships between the variables are studied simultaneously (Brereton 2009). This mathematical process aims to reduce the complexity of the data, identify the inherent trend of a set of experimental measurements and classify the samples according to their similarities and differences, providing graphical visualization in the space defined by the principal components. The principal components were displayed as a set of scores (*t*), which highlights clustering or outliers, and a set of loadings (*p*), which highlights the influence of input variables on *t* (Jackson 2004).

HCA is also an unsupervised method that divides a set of objects into a series of classes to generate a dendrogram in which objects are grouped according to their similarities. There are different criteria to determine the similarity based on correlations (Pearson or otherwise) or distances (Euclidean, Mahalanobis, etc.) (Ramis Ramos and García Álvarez-Coque 2001).

PLS-DA is a supervised model that uses information about the identity of each group of samples to generate a mathematical model that optimizes the separation between classes and allows the class to which new samples belong to be predicted. In this model, PLS was used to find the fundamental relationships between the independent or predictor variables, *X* (in our case, the experimental measurements), and the dependent or response variables, *Y* (in our case, the classes). The dependent

variables ( $Y$  matrix) are categorical and express the fact of belonging or not to a particular class. The matrix consists of as many columns as there are classes, and an observation takes a value of one for the class to which it belongs and zero for the rest (Eriksson et al. 2006; Sjoström et al. 1986). PLS-DA provides rugged models with predictive capability that allows the differences between two sample groups to be found.

Validation of data models is essential to check whether the conclusions obtained are reliable according to the information provided by the model. Validation of the models was tested by internal validation, also called “scrambling”. For this, the parameters  $R^2$  (total variance explained by the model) and  $Q^2$  (predictive capability of the model) were evaluated. The idea of this validation is to compare the values of  $R^2$  and  $Q^2$  of the initial model with the values of the models generated after randomly permuting (about 100 times) the observations of the matrix of the dependent variables while the matrix of independent variables remains intact. The validation is successful if the models obtained after the permutations have  $R^2$  and  $Q^2$  values lower than those of the original model; moreover, the regression line connecting the values of  $Q^2$  intersects the  $Y$ -axis below 0.05 (Eriksson et al. 2006).

## Results and Discussion

### Analysis of Isoflavones

To study the isoflavone content of chickpeas, lentils and beans, 20 different samples of each type of legume were analysed in triplicate. Estimation of isoflavone contents was performed with the external standard calibration method.

The isoflavone contents showed great variability, depending on the type of legume analysed and even among the different varieties from the same species. As an example, the isoflavone compositions of the lentils analysed are shown in Table 1. The data are expressed as means  $\pm$  standard deviations ( $\mu\text{g}/100\text{ g}$ ).

This variability in the isoflavone content not only was manifested in different types of legumes but was also evident in the different varieties of the same species. Differences were observed both in the type of analyte and in the concentration found in the samples. In the three types of legume, the aglycone content found was higher than that of the glucosides.

### Principal Component Analysis

Before applying the multivariate analysis techniques described above, the data were scaled using the Pareto mode.

Using scaled variables, we constructed PCA models capable of classifying (among species, subspecies, varieties or ecotypes) each type of legume sample based on its isoflavone content.

Figure 1 shows the  $PC2 \times PC1$  score plot resulting from the application of PCA to the scaled data for chickpeas, lentils and beans, respectively. As mentioned above, PCA is an unsupervised technique in which the classes are not assigned before the generation of the model. After model construction, each of the scores was coloured according to the label of the producer.

First, the PCA model corresponding to chickpea samples was obtained. When attempting to discriminate between chickpeas according to the manufacturers' labels, no clear differentiation was obtained. It should be noted that the samples labelled by the producer as Maragato, Gordo or Pico Pardal were similar to the Pedrosillano ecotype, both in isoflavone content and morphology; accordingly, these samples could be included in this ecotype. Thus, a new PCA model was generated taking into account only the three ecotypes selected for this study (Castellano, Pedrosillano and Blanco Lechoso). As can be seen in Fig. 1a, as in the previous model, it was not possible to discriminate among the different ecotypes of the same variety of chickpeas (var. *macrocarpum* of *C. arietinum* species).

In the case of lentil samples, as in the above case, the PCA model generated, considering different classes in the product labelling, did not allow classes to be discriminated. In this case, three different varieties of the same species (*L. culinaris* Medicus) were analysed: *vulgaris*, *variabilis* and *dupuyensis*. There were several samples whose labelling did not contain the variety of lentil to which they belonged. To assign them a variety, we considered the percentages of participation of the major analytes (genistein and formononetin). Figure 1b shows the PCA model generated for lentil samples as a function of the three different varieties: *vulgaris*, *variabilis* and *dupuyensis*. As can be seen, PCA allowed a differentiation between the different varieties of lentils, although the class separation was not complete.

Finally, the classification of the bean samples is shown in Fig. 1c. As may be seen, two perfectly differentiated groups were found, corresponding to the two different species studied, *P. vulgaris* and *P. multiflorus*. However, the PCA model did not allow bean subspecies to be differentiated.

### Hierarchical Cluster Analysis

Figure 2 shows the dendrograms generated from the optimum PCA models obtained. For chickpea samples, three groups (or classes) corresponding to the three ecotypes analysed (Blanco Lechoso, Castellano and Pedrosillano) were observed (Fig. 2a). In the study of lentil samples (Fig. 2b), a classification in four groups is shown. This highlights the separation not only between varieties but also between subvarieties. It would be expected that the branch corresponding to the variety *L. culinaris* Medicus var. *vulgaris* would be divided, leading to the subvarieties Rubia de la Armuña and Rubia Castellana. However, these subvarieties showed a greater degree of

**Table 1** Isoflavone content of lentil samples ( $\mu\text{g}/100\text{ g}$ )

	Daidzin	Glycitin	Genistin	Daidzein	Glycitein	Genistein	Formononetin	Biochanin-A
AL	1.32 $\pm$ 0.06	0.79 $\pm$ 0.01	0.51 $\pm$ 0.04	3.2 $\pm$ 0.4	4.4 $\pm$ 0.8	27 $\pm$ 1	15.2 $\pm$ 0.6	18 $\pm$ 1
AL D	nd	nd	nd	0.12 $\pm$ 0.04	nd	5.1 $\pm$ 0.2	0.9 $\pm$ 0.2	nd
AC H	nd	nd	nd	nd	nd	3 $\pm$ 0.2	nd	1.2 $\pm$ 0.4
LI	nd	nd	nd	nd	nd	2.4 $\pm$ 0.2	nd	nd
PS	nd	nd	nd	nd	nd	1.6 $\pm$ 0.9	nd	nd
PL	0.17 $\pm$ 0.02	nd	nd	0.3 $\pm$ 0.1	nd	3.9 $\pm$ 0.9	0.9 $\pm$ 0.2	3 $\pm$ 0.2
PV	0.43 $\pm$ 0.02	0.47 $\pm$ 0.02	0.25 $\pm$ 0.06	nd	nd	7 $\pm$ 0.4	1.9 $\pm$ 0.1	2.1 $\pm$ 0.9
PCI	nd	nd	nd	nd	nd	1.9 $\pm$ 0.2	nd	0.8 $\pm$ 0.4
PBA	nd	nd	0.09 $\pm$ 0.01	nd	nd	2.9 $\pm$ 0.2	0.22 $\pm$ 0.02	1 $\pm$ 0.4
PA	nd	nd	nd	nd	nd	2.4 $\pm$ 0.2	0.18 $\pm$ 0.02	1.1 $\pm$ 0.4
PH	nd	nd	nd	nd	nd	1.6 $\pm$ 0.4	nd	1.8 $\pm$ 0.9
CA	nd	0.11 $\pm$ 0.01	0.18 $\pm$ 0.01	nd	nd	4.0 $\pm$ 0.2	1.17 $\pm$ 0.06	1.9 $\pm$ 0.4
CV	0.086 $\pm$ 0.006	0.1 $\pm$ 0.01	0.18 $\pm$ 0.01	2.52 $\pm$ 0.09	nd	10.4 $\pm$ 0.4	4.64 $\pm$ 0.09	2.3 $\pm$ 0.4
CCU	nd	nd	0.28 $\pm$ 0.01	nd	nd	2.6 $\pm$ 0.2	0.82 $\pm$ 0.06	3.1 $\pm$ 0.4
VV	0.25 $\pm$ 0.02	0.06 $\pm$ 0.01	0.11 $\pm$ 0.06	nd	nd	32.3 $\pm$ 0.9	8.4 $\pm$ 0.2	4.6 $\pm$ 0.9
EA	nd	nd	nd	nd	nd	4.4 $\pm$ 0.2	0.31 $\pm$ 0.06	nd
RH	nd	nd	nd	nd	nd	4.5 $\pm$ 0.2	0.38 $\pm$ 0.06	0.9 $\pm$ 0.4
EC	nd	0.15 $\pm$ 0.004	nd	nd	nd	7.4 $\pm$ 0.2	0.87 $\pm$ 0.06	1.1 $\pm$ 0.4
LH	0.15 $\pm$ 0.02	0.08 $\pm$ 0.01	0.28 $\pm$ 0.06	nd	nd	3.1 $\pm$ 0.4	2.6 $\pm$ 0.1	1.9 $\pm$ 0.9
LP	nd	nd	0.19 $\pm$ 0.02	0.28 $\pm$ 0.06	nd	3.4 $\pm$ 0.2	3.1 $\pm$ 0.2	nd

differentiation, which could be explained in terms of the contents of the major analyte (formononetin). The formononetin content was higher in the scores corresponding to the subvariety Rubia de la Armuña than in those of other varieties and subvarieties. This higher content of formononetin could be explained by taking into account the origin of the samples; such behaviour has been described previously in the literature (Morrison et al. 2010).

Figure 2c-1 shows the dendrogram obtained in the case of bean samples, in which the separation of the two species (*P. vulgaris* and *P. multiflorus*) was clearly apparent. The figure shows that the group 2 (*P. vulgaris*) was subdivided into two distinct groups, which could correspond to the two subspecies of *P. vulgaris* studied, *nanus* and *volubilis*. Figure 2c-2 shows the dendrogram obtained considering both species and subspecies. In this figure, groups 3 and 4 represented the subspecies *nanus* and *nolubilis*, respectively. Group 2 corresponded to samples labelled by the supplier as “Riñón” beans, which belonged to the subspecies *nanus*; however, they were very different from these, both morphologically and considering their isoflavone contents. This prompted us to suspect that they were a different kind of bean.

### Partial Least Squares Discriminant Analysis

Finally, partial least squares discriminant analysis (a supervised technique) was conducted. This analysis allowed a more

effective separation between classes, since the variables affecting the separation were known. The PLS-DA models (Fig. 3) were generated from dendrograms obtained previously. For chickpea samples, PCA models did not allow the different ecotypes to be differentiated. However, as in HCA, PLS-DA showed a clear separation between Blanco Lechoso and the other two ecotypes, Castellano and Pedrosillano (Fig. 3a).

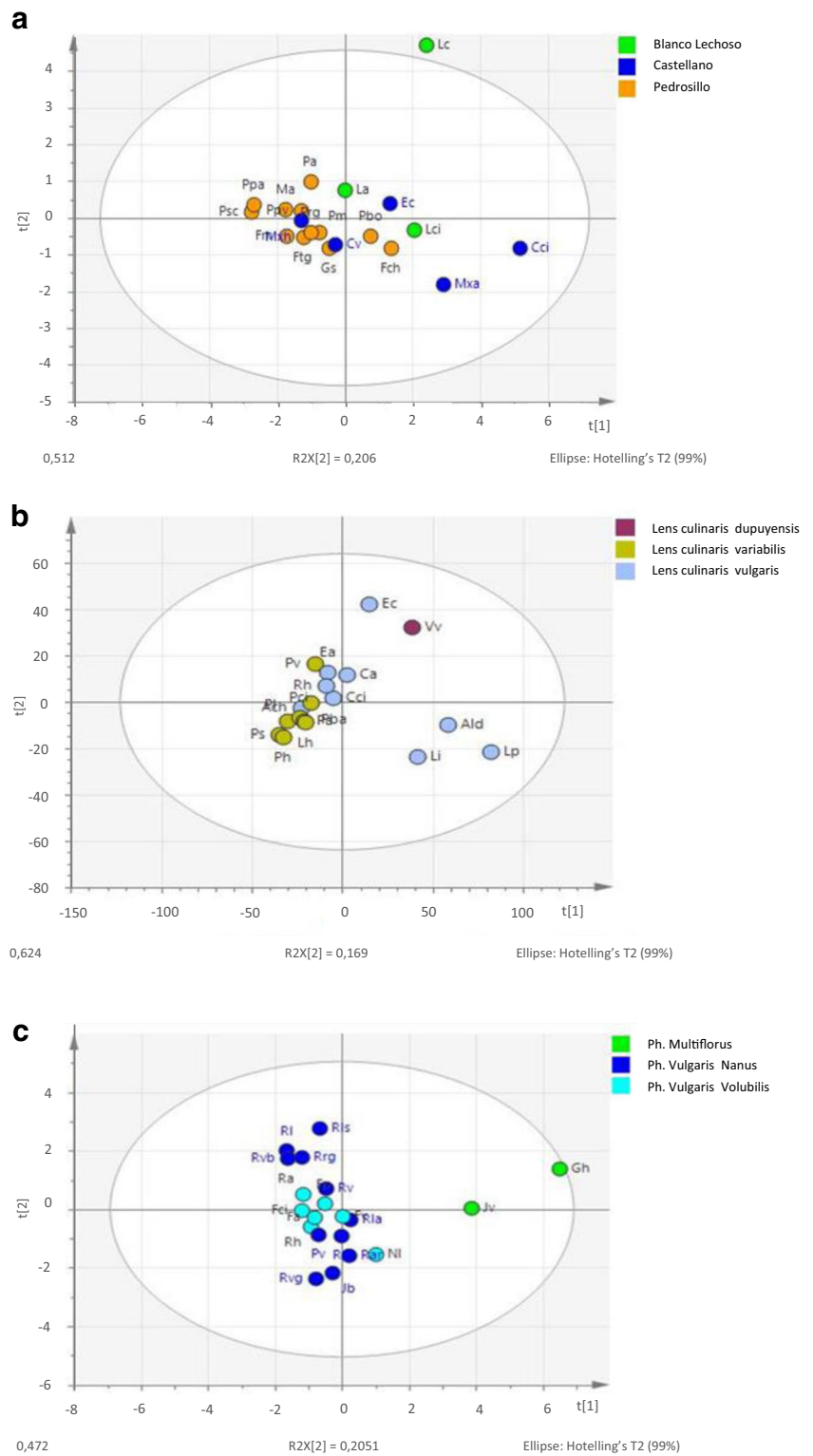
In the case of lentil samples, the results obtained with PLS-DA corroborated the conclusions reached with HCA. Figure 3b shows the discrimination between four groups corresponding to the different varieties and subvarieties. As in HCA analysis, the separation between Rubia de la Armuña lentils and the other samples was greater than that obtained for the rest of the groups, which could be explained by the isoflavone content as well as the origin of the samples.

Finally, PLS-DA was applied to the bean samples. Figure 3c shows a clear separation between two species: *P. multiflorus* (group 1) and *P. vulgaris* (groups 2, 3 and 4). As in HCA, when attempting to differentiate between the two subspecies of *P. vulgaris*, a group of samples (group 2) appeared separated from the group to which it should belong (group 3).

### Method Validation

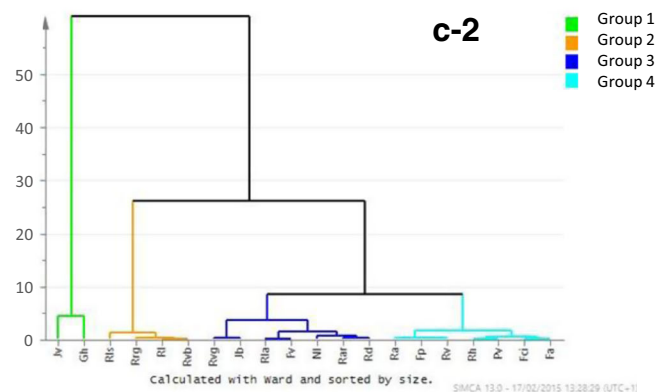
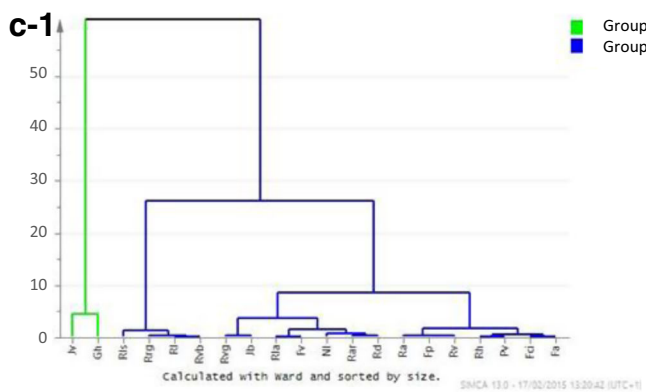
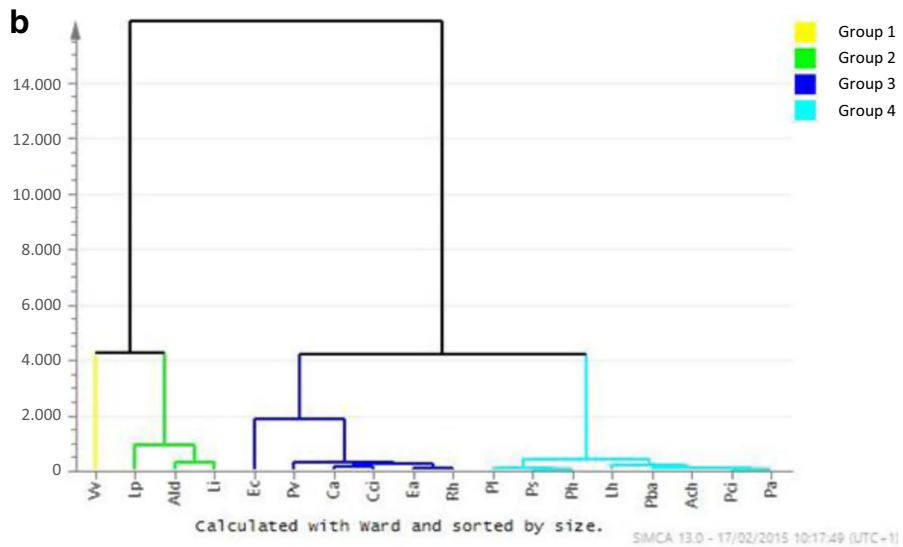
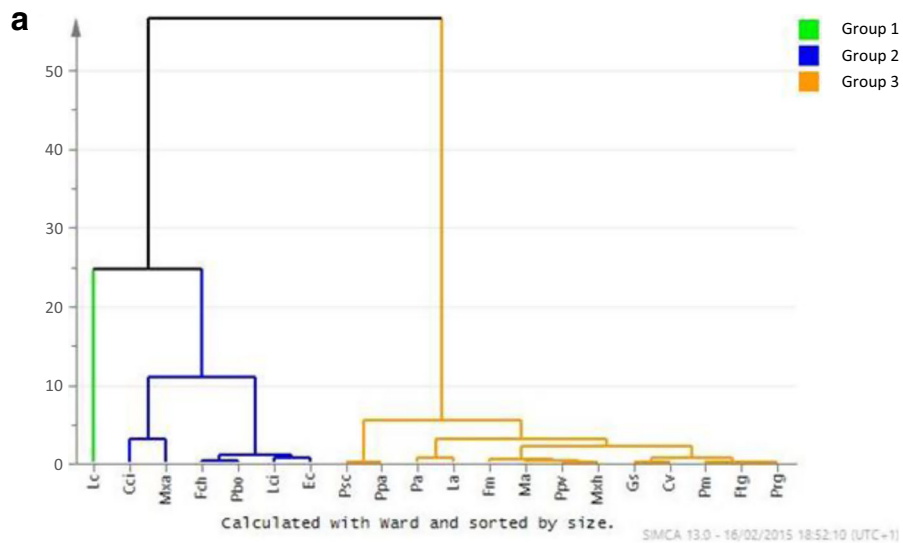
After generating the models, they were validated to check the reliability of the conclusions based on the information provided by them. For this, the parameters R<sup>2</sup> (total variance explained

**Fig. 1** Principal component analysis of the isoflavone contents of legume samples. **a** Chickpeas: samples denominated by their ecotypes. **b** Lentils: samples denominated by their varieties. **c** Beans: samples denominated by their species and subspecies



by the model) and Q2 (predictive capability of the model) were evaluated with the permutation method called scrambling validation. As mentioned above, the validation is successful if the values of R2 and Q2 of the models after the permutations are lower than those of the original model and the regression line

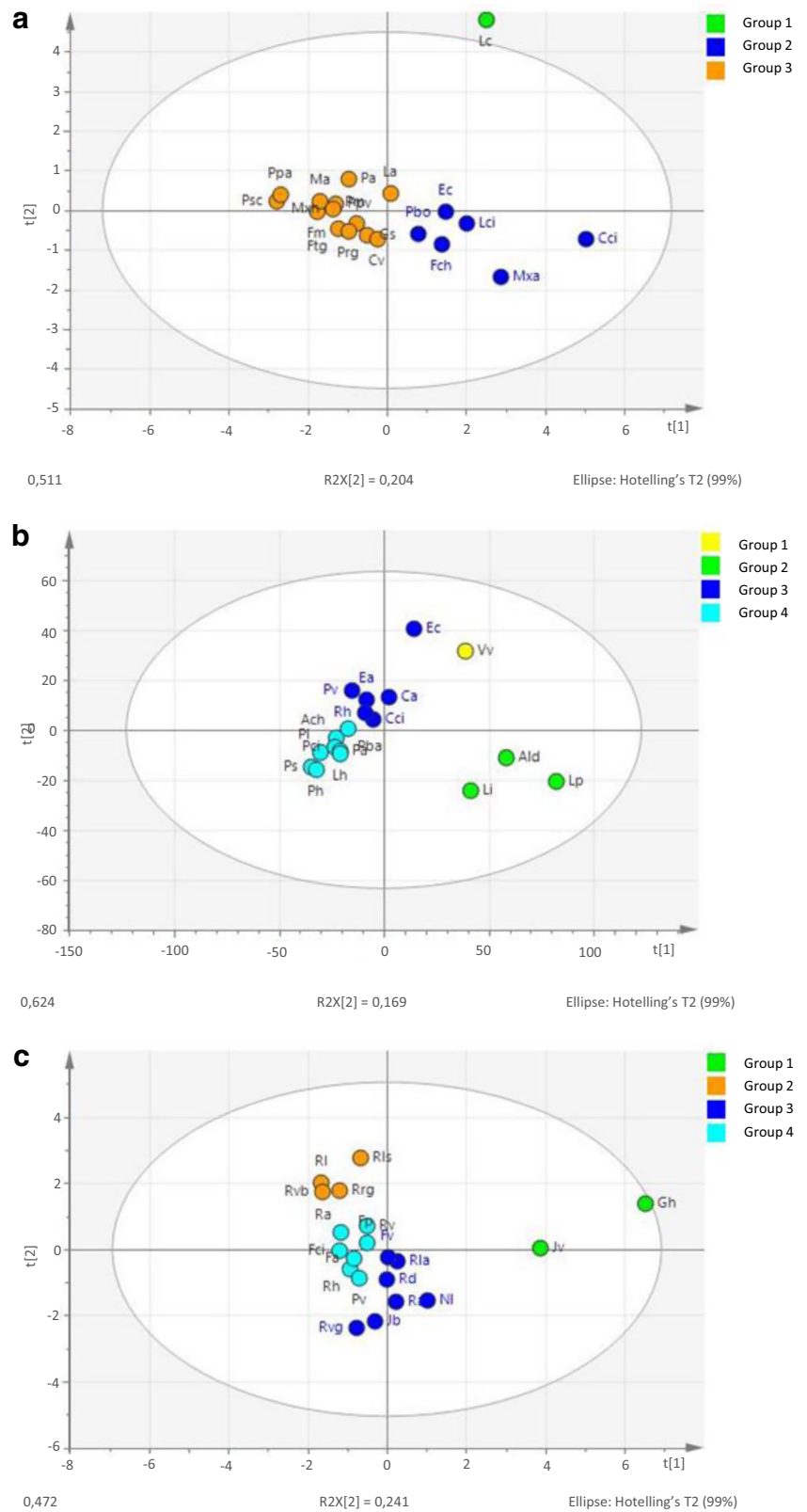
connecting the values of Q2 intersects the Y-axis below 0.05. The upper right part of Fig. 4 shows the values of R2 and Q2 of the original model, while on the left, the figure shows the values of R2 and Q2 of the models after the permutations. As can be seen, the conditions required were fulfilled, taking the intercept



**Fig. 2** Hierarchical cluster analysis of isoflavone contents of legume samples. **a** Chickpeas: group 1 “Blanco Lechoso”, group 2 “Castellano”, and group 3 “Pedrosillano”. **b** Lentils: group 1 “Verdina”, group 2 “Rubia de la Armaña”, group 3 “Rubia

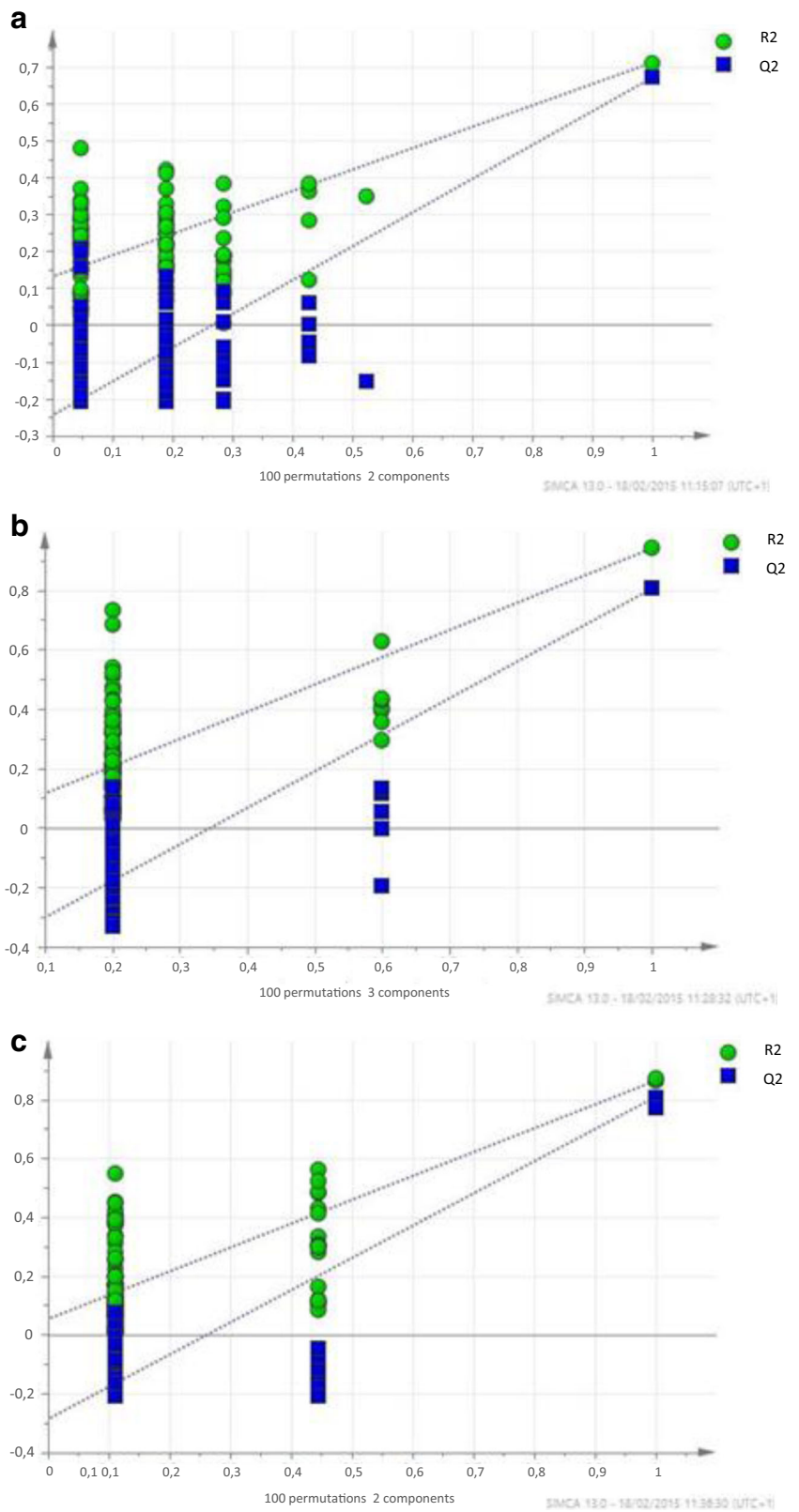
Castellana”, and group 4 “Pardina”. **c** Beans: **c-1** group 1 “*Phaseolus multiflorus*”, group 2 “*Phaseolus vulgaris*”; **c-2** group 1 “*Phaseolus multiflorus*”, group 2 “unknown”, group 3 “*P. vulgaris* spp. *nanus*”, and group 4 “*P. vulgaris* spp. *volubilis*”

**Fig. 3** Partial least squares discriminant analysis of isoflavone contents of legume samples. **a** Chickpeas: group 1 “Blanco Lechoso”, group 2 “Castellano”, and group 3 “Pedrosillano”. **b** Lentils: group 1 “Verdina”, group 2 “Rubia de la Arnuña”, group 3 “Rubia Castellana”, and group 4 “Pardina”. **c** Beans: group 1 “*Phaseolus multiflorus*”, group 2 “unknown”, group 3 “*Phaseolus vulgaris* spp. *nanus*”, group 4 “*Phaseolus vulgaris* spp. *volubilis*”





**Fig. 4** Scrambling validation of the model. **a** Chickpeas. **b** Lentils. **c** Beans



of the Q2 line values of  $-0.245$ ,  $-0.425$  and  $-0.287$  for chickpea, lentil and bean samples, respectively.

## Conclusions

In this study, the isoflavone profiles of different types of chickpeas, lentils and beans were obtained and evaluated. The legumes analysed exhibited different isoflavone profiles. In addition, within the same genus of legumes, there are differences in the isoflavone content between species, subspecies, varieties and even ecotypes of the considered legume.

To study the possibility of classifying the different types of pulses as a function of their isoflavone content, multivariate analysis techniques were applied.

According to the obtained results, it can be concluded that, using generated PCA models only, the different species of beans could be distinguished, but not subspecies.

The results provided by HCA revealed that the different species and subspecies of beans and different varieties and subvarieties of lentils can be distinguished, and—in the case of chickpeas—even the different ecotypes of the same variety. It should be noted that in this classification, not only the genotype of the species was taken into account but also the origin of the samples. The results provided by PLS-DA corroborated the above.

The relevance of this study is clear since the differences between pulses of a given variety based on their isoflavone content are corroborated. This study is important for the quality authentication of legumes, as it can allow for both the classification of the different kinds of legumes as well as the detection of fraud by adulteration.

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

**Conflict of Interest** Lara Pérez-Martín declares that she has no conflict of interest. Myriam Bustamante-Rangel declares that she has no conflict of interest. M. M. Delgado-Zamarreño declares that she has no conflict of interest.

**Informed Consent** Not applicable.

**Funding** This study was funded by the “Programa Propio de Financiación de Grupos de Investigación de la Universidad de Salamanca” Ref: USAL2015-16.

## References

- Adlercreutz H (1998) Epidemiology of phytoestrogens. *Baillière's Clin Endocrinol Metab* 12:605–623
- Brereton R (2009) *Chemometrics for pattern recognition*. Wiley, New York, pp. 47–74
- Bustamante-Rangel M, Pérez-Martín L, Delgado-Zamarreño MM (2014) Comparative study of the methodology used in the extraction of isoflavones from legumes applying a modified QuEChERS approach. *Phytochem Anal* 25:170–177
- Cano A, García-Pérez MA, Tarín JJ (2010) Isoflavones and cardiovascular disease. *Maturitas* 67:219–226
- Cassidy A, Albertazzi P, Nielsen IL, Hall W, Williamson G, Tetens I, Atkins S, Cross H, Manios Y, Wolk A, Steiner C, Branca F (2006) Critical review of health effects of soybean phyto-oestrogens in post-menopausal women. *Proc Nutr Soc* 65:76–92
- Delgado-Zamarreño MM, Pérez-Martín L, Bustamante-Rangel M, Carabias-Martínez R (2012) A modified QuEChERS method as sample treatment before the determination of isoflavones in foods by ultra-performance liquid chromatography–triple quadrupole mass spectrometry. *Talanta* 100:320–328
- Eriksson L, Johansson E, Kettaneh-Wold N, Trygg J, Wikström C, Wold S (2006) Multi- and megavariate data analysis. Part 1: basic principles and applications. *Umetrics Academy*, Umea, pp 361–380.
- Frederick M, McCarty MF (2006) Isoflavones made simple—Genistein's agonist activity for the beta-type estrogen receptor mediates their health benefits. *Med Hypotheses* 66:1093–1114
- Hale GE, Hughes CL, Cline JM (2002) Endometrial cancer: hormonal factors, the perimenopausal “window of risk,” and isoflavones. *J Clin Endocrinol Metab* 87:3–15
- Hwang CS, Kwak HS, Lim HJ, Lee SH, Kang YS, Choe TB, Hur HG, Han KO (2006) Isoflavone metabolites and their in vitro dual functions: they can act as an estrogenic agonist or antagonist depending on the estrogen concentration. *J Steroid Biochem Mol Biol* 101:246–253
- Jackson JE (2004) *A user's guide to principal components*. Wiley, New York, pp. 4–25
- Jacobs A, Wegewitz U, Sommerfeld C, Grossklaus LA (2009) Efficacy of isoflavones in relieving vasomotor menopausal symptoms—a systematic review. *Mol Nutr Food Res* 53:1084–1097
- Kumar V, Rani A, Dixit AK, Pratap D, Bhatnagar D (2010) A comparative assessment of total phenolic content, ferric reducing-antioxidative power, free radical-scavenging activity, vitamin C and isoflavones content in soybean with varying seed coat colour. *Food Res Int* 43:323–328
- Lee SJ, Seguin P, Kim JJ, Moond HI, Ro HM, Kim EH, Seo SH, Kang EY, Ahn JK, Chung IM (2010) Isoflavones in Korean soybeans differing in seed coat and cotyledon color. *J Food Compos Anal* 23:160–165
- Martin LM, Castilho MC, Silveira MI, Abreu JM (2006) Liquid chromatographic validation of a quantitation method for phytoestrogens, biochanin-A, coumestrol, daidzein, formononetin, and genistein, in lucerne. *J Liq Chromatogr Rel Technol* 29:2875–2884
- Mazur WM, Duke JA, Wähälä K, Rasku S, Adlercreutz H (1998) Isoflavonoids and lignans in legumes: nutritional and health aspects in humans. *J Nutr Biochem* 9:193–200
- Morris C, Thorpe J, Ambrosio L, Santin M (2006) The soybean isoflavone genistein induces differentiation of MG63 human osteosarcoma osteoblasts. *J Nutr* 136:1166–1170
- Morrison MJ, Cober ER, Saleem MF, McLaughlin NB, Frégeau-Reid J, Ma BL, Woodrow L (2010) Seasonal changes in temperature and precipitation influence isoflavone concentration in short-season soybean. *Field Crops Res* 117:113–121
- Ramis Ramos G, García Álvarez-Coque C (2001) *Quimiometría*. Síntesis S.A., Madrid.
- Raschke M, Wähälä K, Pool-Zobel BL (2006) Reduced isoflavone metabolites formed by the human gut microflora suppress growth but do not affect DNA integrity of human prostate cancer cells. *J Nutr* 96:426–434
- Rimbach G, Boesch-Saadatmandi C, Frank J, Fuchs D, Wenzel U, Daniel H, Hall WL, Weinberg PD (2008) Dietary isoflavones in the

- prevention of cardiovascular disease—a molecular perspective. *Food Chem Toxicol* 46:1308–1319
- Sjostrom M, Wold S, Soderstrom B (1986) PLS discriminant plots. In: *Pattern recognition in practice II*. Elsevier, Amsterdam, p. 486
- Trock BJ, Hilakivi-Clarke L, Clarke R (2006) Meta-analysis of soy intake and breast cancer risk. *J Natl Cancer I* 98:459–471
- Vamerali T, Barion G, Hewidy M, Mosca G (2012) Soybean isoflavone patterns in main stem and branches as affected by water and nitrogen supply. *Eur J Agron* 41:1–10
- Wood CE, Register TC, Franke AA, Anthony MS, Cline JM (2006) Dietary soy isoflavones inhibit estrogen effects in the postmenopausal breast. *Cancer Res* 66:1241–1249