Original papers

Phenotypic identification of plum varieties (*Prunus domestica* L.) by endocarps morpho-colorimetric and textural descriptors

Marco Sarigu[a, b, c]

Oscar Grillo[a, b, c, d, *]

oscar.grillo.mail@gmail.com

Marisol Lo Bianco[d]

Mariano Ucchesu[a, b, c]

Guy d'Hallewin[e]

Maria Cecilia Loi[c]

Gianfranco Venora[d]

Gianluigi Bacchetta[a, b, c]

[a]Banca del Germoplasma della Sardegna (BG-SAR), Hortus Botanicus Karalitanus (HBK), Università degli Studi di Cagliari, Viale Sant'Ignazio da Laconi, 9-11, 09123 Cagliari, Italy

[b]Centro Conservazione Biodiversità (CCB), Dipartimento di Scienze della Vita e dell'Ambiente (DISVA), Università degli Studi di Cagliari, Viale Sant'Ignazio da Laconi, 11-13, 09123 Cagliari, Italy

[c]Dipartimento di Scienze della Vita e dell'Ambiente (DISVA), Viale Sant'Ignazio da Laconi, 13, 09123 Cagliari, Italy

[d]Stazione Consorziale Sperimentale di Granicoltura per la Sicilia, Via Sirio 1, 95041 Borgo Santo Pietro, Caltagirone, Italy

[e]Consiglio Nazionale delle Ricerche, Istituto di Scienze delle Produzioni Alimentari, Traversa La Crucca 3, Loc. Baldinca, 07040 Li Punti, Sassari, Italy

*Corresponding author at: Stazione Consorziale Sperimentale di Granicoltura per la Sicilia, Via Sirio 1, 95041 Borgo Santo Pietro, Caltagirone, Italy.

Abstract

The identification of plum varieties is generally done on the base of distinctive plant traits such as shape, size, and fruit drupe color identified during the variety registration, following official descriptors. In this paper, image analysis techniques were applied to study endocarps variability of 23 *Prunus domestica* cultivars from Sardinia. Digital images were acquired and analysed using a macro specifically developed to measure morpho-colorimetric endocarps features. The data were later statistically processed applying the stepwise Linear Discriminant Analysis (LDA) to implement a statistical classifier able to classify each variety and identify plausible synonymy groups.

The present study represent the first attempt to investigate the morphology and morphometry of plum endocarps in order to characterize the whole Sardinian plum agrobiodiversity. It is also the evidence of the usefulness of image analysis techniques in taxonomic investigations too, as well as for the conservation and enhancement of traditional plums for consumer satisfaction.

Keywords: Elliptical Fourier shape Descriptors (EFDs); Endocarp characterization; Haralick descriptors; Image analysis; *Prunus domestica* L.

1 Introduction

Plum (*P. domestica* L.), together with apple and pear, is one of the most important fruits in temperate regions of the world (Zohary et al., 2012). Plums are grouped into three main categories, *Prunus salicina* Lindl., *Prunus domestica* L., and [*Prunus domestica* subsp. *insititia* (L.) Bonnier & Layens], cultivated before the introduction of *P. domestica*. Many authors are agree on the hypothesis that *P. domestica* derived from one or

a combination of several Eurasian progenitors as *P. domestica* subsp. *insititia*, *P. cerasifera* Ehrh. and *P. spinosa* L. (Faust and Suranyi, 1999; Zohary et al., 2012).

Different studies claim that *P. spinosa,* is a Crop Wild Relative (CWR) with a secondary gene pool that have contributed to generate the domesticated form of plum, although the two species are morphologically distinct (~~Eryomine, 1990,~~ Zohary et al., 2012). Recent genetic studies have shown that *P. spinosa* and *P. domestica* subsp. *insititia* have close relationships with the current European domestic plums (Pollmann et al., 2005; Horvath et al., 2011; Athanasiadis et al., 2013).

P. domestica seems to have originated in Southern Europe or Western Asia between the Caucasus Mountains and the Caspian Sea overlapping with the centre of origin of *P. cerasifera* and from there moved into Western Europe.

The earliest archaeological remains of *P. domestica* in Europe are attributable to the Roman Period (Faust and Suranyi, 1999; Feemster Jashemski and Meyer, 2002; Zohary et al., 2012). During this Period, the domestic plum seems to appear and spread in western Europe and for this reason plum seem to appear mostly in the Roman waterlogged archaeological context (Janick, 2005).

P. domestica fruits exhibit a great diversity in size, shape, color and taste. There are significant differences in color among plum fruits. Usenik et al. (2009) demonstrated that there is a wide range of variability in the anthocyanins and chromatic parameters during fruit ripening of *P. domestica* varieties.

The endocarp can be smooth or wrinkled, round, globular or elliptical and it can be sharp or smooth with slightly conical base. The dorsal and lateral surfaces of the endocarp can also present crests (Agabbio et al., 2016).

Commonly, variety identification is done following official descriptors, based on morphological and physiological characters of the plant and until some years ago, the dimensional measurements of endocarps were done manually, generally by calipers, based on fixed categories officially recognized (Horvath et al., 2011). For this reason, in literature there are only studies on *P. domestica* endocarps based on classical morphologic and morphometric techniques (Nielsen and Olrik, 2001; Pollmann et al., 2005; Depypere et al., 2007).

In the last two decades, a significant increase in image analysis applications has been highlighted in the plant biology research field and automatized system have the potential to replace human visual assessments. Many recent papers testify the importance of the biometric features, measured by computer vision techniques, in taxonomic studies, to characterize and identify wild plant species (Rovner and Gyulai, 2007; Fawzi, 2011; Grillo et al., 2012, 2013; Pinna et al., 2014; Lo Bianco et al., 2015a; Santo et al., 2015). This has stimulated research in many areas, including the agronomical field (Grillo et al., 2011; Orrù et al., 2012, 2015; Smykalova et al., 2011, 2013; Lo Bianco et al., 2015b; Sabato et al., 2015; Gresta et al., 2016; Piras et al., 2016). Also, this methodology was applied in the archaeobotanical field (Bouby et al., 2013; Ucchesu et al., 2015, 2016).

In this view, the aim of this study is to measure endocarp descriptive features such as shape, size, surface color and texture by computer image analysis in order to:

- ✓ build a database of phenotypic features of *P. domestica* endocarps;
- ✓ implement, on the basis of the developed database, a statistical classifier able to identify and classify each variety of *P. domestica*;
- ✓ carry out hypothetical synonymy groups.

This study represents the first morpho-colorimetric approach on *P. domestica* L. varieties of Sardinia based on endocarp image analysis.

2 Material and methods

2.1 Plant ~~M~~material

Fruits of *P. domestica* (1663 endocarps) from 23 traditional varieties of Sardinia were collected from the field catalogue of CNR-ISPA (Nuraxinieddu, OR, Sardinia) (Table 1). Endocarps of *P. spinosa* (984 samples) were collected from 11 populations of Sardinia (Italy) and they were used for comparison with the cultivated one (data not shown). In additions, two commercial varieties, Mirabolano Giallo (MIB) and Mirabolano Rosso (MIR), were included in this study as outgroup, because phenotypically similar and the most representative at national level. The sampling procedure provided the collection of one to five trees for each variety, randomly collecting mature fruits at the time of the maximum concentration of sugar in the pulp (in summer). In order to reduce the environmental effects, the fruit sampling was conducted for three years. Exocarps and mesocarps of each fruit were manually removed, and the endocarps accurately cleaned, washed and dried.

Table 1 General information about *P. domestica* varieties from the field catalogue of CNR-ISPA.

Sample code	Cultivar	Sampling location	Endocarps amount
BOS	Bosana di Agosto	Bosa	72
CAD	Cariadoggia	Alghero	80
CAR	Cariasina	Medio Campidano	39
COR	Coru	Laconi	85
CCO	Coru ‘e Columbu	Laconi	80
CRO	Croccorighedda	Laconi	85
DOA	Dore A	Alghero	30
BON	Di Bonarcado	Bonarcado	100
GIB	Gialla di Bosa	Bosa	60
LUI	Luisa	Laconi	30
LA1	Laconi Rosata	Laconi	90
LA2	Groga	Laconi	90
LA3	Ollanu de Ou	Laconi	85
LA4	Laconi E	Laconi	30
LA5	Laconi F	Laconi	70
MEL	Meloni	Gonnosfanadiga	90
LIM	Limuninca	Sassari	60
NES	Nera Sarda	Bosa	100
SAG	San Giovanni	Oristano	57
SB1	Sanguigna di Bosa	Bosa	90
SB2	Sanguigna di Bosa	Bosa	60
SAE	Sant'Elia	Nuoro	90
SIG	Sighera	Gonnosfanadiga	90
MIG	Mirabolano Giallo	Commercial	90
MIB	Mirabolano Rosso	Commercial	90

Later the endocarps of all varieties were preserved in the Sardinian Germoplasm Bank (BG-SAR) of the University of Cagliari (Atzeri et al., 2012).

2.2 Image analysis

Digital images of endocarps were acquired using a flatbed scanner (Epson Perfection V550 Photo), with a digital resolution of 400 dpi. The images were processed and analysed using the software package KS-400 V. 3.0. (Carl Zeiss, Vision, Oberkochen, Germany). According to Shahin and Symons (2003), before the acquisition of the sample images, the scanner was calibrated for color matching, using a Q60 Kodak Color Input Target chart.

A macro, specifically developed for the characterization of wild seeds (Bacchetta et al., 2008) and later modified to measure a further 20 morpho-colorimetric seed features (Mattana et al., 2008), was adapted to automatically perform the whole analysis procedure, reducing the execution time and contextual mistakes in the analysis process (Grillo et al., 2010). This macro, called *Prunus*.mcr, was further enhanced adding algorithms able to compute the Elliptic Fourier Descriptors (EFDs) for each analysed endocarp, increasing the number of discriminant parameters (Table 2, Fig. 1).

Table 2 List of morphometric features measured on endocarps, calculated according to Hâruta (2011), excluding the Elliptic Fourier Descriptors (EFDs) and the Haralick's descriptors reported in Table 3.

	Feature	Description
A	Area	Endocarp area (mm^2)
P	Perimeter	Endocarp perimeter (mm)
P_{conv}	Convex Perimeter	Convex perimeter of the endocarp (mm)
P_{Crof}	Crofton Perimeter	Crofton perimeter of the endocarp (mm)
P_{conv}/P_{Crof}	Perimeter ratio	Ratio between convex and Crofton's perimeters
D_{max}	Max diameter	Maximum diameter of the endocarp (mm)
D_{min}	Min diameter	Minimum diameter of the endocarp (mm)
D_{min}/D_{max}	Feret ratio	Ratio between minimum and maximum diameters
Sf	Shape Factor	Endocarp shape descriptor = $(4 \times \pi \times \text{area})/\text{perimeter}^2$ (normalized value)
Rf	Roundness Factor	Endocarp roundness descriptor = $(4 \times \text{area})/(\pi \times \text{max diameter}^2)$ (normalized value)
Ecd	Eq. circular diameter	Diameter of a circle with equivalent area (mm)
F	Fiberlength	Endocarp length along the fiber axis
C	Curl degree	Ratio between D_{max} and F
$Conv$	Convessity degree	Ratio between P_{Crof} and P
Sol	Solidity degree	Ratio between A and convex area
Com	Compactness degree	Endocarp compactness descriptor = $[\sqrt{(4/\pi)A}]/D_{max}$
EA_{max}	Maximum ellipse axis	Maximum axis of an ellipse with equivalent area (mm)
EA_{min}	Minimum ellipse axis	Minimum axis of an ellipse with equivalent area (mm)
R_{mean}	Mean red channel	Red channel mean value of endocarp pixels (grey levels)
R_{sd}	Red std. deviation	Red channel standard deviation of endocarp pixels
G_{mean}	Mean green channel	Green channel mean value of endocarp pixels (grey levels)
G_{sd}	Green std. deviation	Green channel standard deviation of endocarp pixels
B_{mean}	Mean blue channel	Blue channel mean value of endocarp pixels (grey levels)
B_{sd}	Blue std. deviation	Blue channel standard deviation of endocarp pixels
H_{mean}	Mean hue channel	Hue channel mean value of endocarp pixels (grey levels)
H_{sd}	Hue std. deviation	Hue channel standard deviation of endocarp pixels

L_{mean}	Mean lightness ch.	Lightness channel mean value of endocarp pixels (grey levels)
L_{sd}	Lightness std. dev.	Lightness channel standard deviation of endocarp pixels
S_{mean}	Mean saturation ch.	Saturation channel mean value of endocarp pixels (grey levels)
S_{sd}	Saturation std. dev.	Saturation channel standard deviation of endocarp pixels
D_{mean}	Mean density	Density channel mean value of endocarp pixels (grey levels)
D_{sd}	Density std. deviation	Density channel standard deviation of endocarp pixels
S	Skewness	Asymmetry degree of intensity values distribution (grey levels)
K	Kurtosis	Peakness degree of intensity values distribution (densit. units)
H	Energy	Measure of the increasing intensity power (densitometric units)
E	Entropy	Dispersion power (bit)
D_{sum}	Sum of Density	Sum of Density values of the endocarp pixels (grey levels)
SqD_{sum}	Sum of the Squares of density	Sum of the Squares of density values (grey levels)

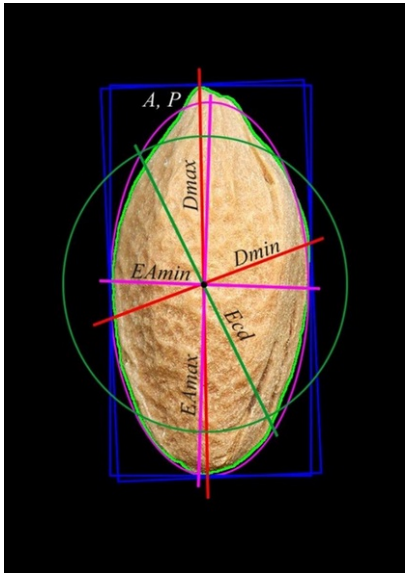

Fig. 1

As described by Lo Bianco et al. (2015a), this method allows describing the boundary of the seed projection, as an array of complex numbers, which correspond to the pixel position of the seed boundary. About the use of number of harmonics for an optimal description of outlines, in order to minimize the measurement errors and optimize the efficiency of shape reconstruction, 20 harmonics were used to define the endocarp boundaries, obtaining further 77 parameters useful to discriminate among the studied varieties of *Prunus*.

Finally, the macro was further improved adding algorithms able to compute 11 Haralick's descriptors and the relative standard deviation values for each analysed endocarp (Table 3). These parameters are generally used to accurately describe the surface texture of an object based on grey tonal features (Haralick et al., 1973; Haralick and Shapiro, 1991). A total of 134 morpho-colorimetric and textural features were measured.

Table 3 Haralick's descriptors measured as reported in Haralick et al. (1973).

	Feature	Equation
Har 1	Angular second moment	$\sum_i \sum_j p(i,j)^2$
Har 2	Contrast	$\sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j) \right\}, \ \lvert i,j \rvert = n$
Har 3	Correlation	$\frac{\sum_i \sum_j (ij) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$
		where μ_x, μ_y, σ_x and σ_y are the means and the standard deviations of p_x and p_y
Har 4	Sum of square: variance	$\sum_i \sum_j (i-\mu)^2 p(i,j)$
Har 5	Inverse difference moment	$\sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$
Har 6	Sum average	$\sum_{n=2}^{2N_g} i p_{x+y}(i)$
		where x and y are the coordinates (row and column) of an entry in the co-occurrence matrix, and $p_{x+y}(i)$ is the probability of co-occurrence matrix coordinates summing to $x + y$
Har 7	Sum variance	$\sum_{i=2}^{2N_g} (i - f_8)^2 p_{x+y}(i)$
Har 8	Sum entropy	$-\sum_{i=2}^{2N_g} p_{x+y}(i) \log\{p_{x+y}(i)\} = f_8$
Har 9	Entropy	$-\sum_i \sum_j p(i,j) \log[p(i,j)]$
Har 10	Difference variance	$\sum_{n=0}^{N_{g-1}} i^2 p_{x-y}(i)$
Har 11	Difference entropy	$-\sum_{n=0}^{N_{g-1}} p_{x-y}(i) \log\{p_{x-y}(i)\}$

The basis for these features is the grey-level co-occurrence matrix (G in equation (1)). This matrix is square with dimension Ng, where Ng is the number of grey levels in the image. Element $[i,j]$ of the matrix is generated by counting the number of times a pixel (p) with value i is adjacent to a pixel with value j and then dividing the entire matrix by the total number of such comparisons made. Each entry is therefore considered the probability that a pixel with value i will be found adjacent to a pixel of value j.

$$G = \begin{bmatrix} p(1,1) & p(1,2) & \cdots & p(1,N_g) \\ p(2,1) & p(2,2) & \cdots & p(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_g,1) & p(N_g,2) & \cdots & p(N_g,N_g) \end{bmatrix} \quad (1)$$

2.3 Data analysis

The achieved results were used to build a database of morpho-colorimetric and texture features. Statistical elaborations were executed using the software IBM SPSS (Statistical Package for Social Science) release 16.0 (SPSS Inc. for Windows, Chicago, Illinois, USA), and the stepwise Linear Discriminant Analysis method (LDA) was applied to compare all the *P. domestica* endocarps.

LDA method is commonly used to classify or identify unknown groups characterized by quantitative and qualitative variables (Fisher, 1936, 1940; Sugiyama, 2007), finding the combination of predictor variables with the aim of minimizing the within-class distance and maximizing the between-class distance simultaneously, thus achieving maximum class discrimination (Hastie et al., 2001; Holden et al., 2011; Kuhn and Johnson, 2013). The stepwise method identifies and selects the most statistically significant features among the 98 measured on each endocarp, using three statistical variables: Tolerance, *F*-to-enter and *F*-to-remove. The Tolerance value indicates the proportion of a variable variance not accounted for by other independent variables in the equation. *F*-to-enter and *F*-to-remove values define the power of each variable in the model and are useful to describe what happens if a variable is inserted and removed, respectively, from the current model. This method starts with a model that does not include any of the variables. At each step, the variable with the largest *F*-to-enter value that exceeds the entry criterion chosen ($F \geq 3.84$) is added to the model. The variables left out of the analysis at the last step have *F*-to-enter values smaller than 3.84, and therefore no more are added. The process is automatically stopped when no remaining variables are able to

increase the discrimination ability (Grillo et al., 2012). Finally, a cross-validation procedure was applied to verify the performance of the identification system, testing individual unknown cases and classifying them based on all others (SPSS, 2006).

3 Results

3.1 Comparison among wild and cultivated Prunus samples

A preliminary comparison was executed between the two groups of studied samples: *P. spinosa* populations and *P. domestica* varieties. An overall percentage of correct identification of 99.3% was reached. *P. spinosa* populations and *P. domestica* varieties reached high performance of correct identification of 99.3% (Table 4).

Table 4 Comparison between cultivated and wild samples of *Prunus*.

	P. domestica L.	*P. spinosa* L.	Total
P. domestica L.	98.8 (1820)	1.1 (23)	100.0 (1843)
P. spinosa L.	0.2 (5)	99.8 (2125)	100.0 (2130)
Overall			99.3% (3973)

3.2 Varieties comparison

Another analysis was implemented in order to compare the studied varieties. From this comparison, an overall correct identification performance of 86.1% was reached. The classifier was able to identify the varieties with matching percentages, ranging between 68.3% for Sanguigna di Bosa 2 (SB2) and 97.5%, for Cariadoggia (CAD), except for Ollanu de Ou (LA3) and Bosana which reached the highest score of correct classification (100%), and the variety Coru e Columbu (CCO) which were correctly identified only in 60.0% of the cases (Table 6).

3.3 Synonym groups

Considering the achieved results, in order to evaluate possible similarities and differences, and contextually identify possible synonymy groups a further comparative analysis was conducted considering all *P. domestica* varieties. Although the effect of all the varieties considered together caused a significant reduction of the identification performance the amplified mutual misidentification between the varieties Coru (COR) and Coru ‘e Columbu (CCO) with percentages of 14.1% and 30.0%, respectively, suggest that these two nominal varieties could be synonyms of the same variety. At the same time, the high mutual misattribution also occurs between the varieties Sanguigna di Bosa 1 (SB1) and Sanguigna di Bosa 2 (SB2) with percentages of 7.8% and 8.3%, respectively.

The commercial samples, Mirabolano Giallo (MIB) and Mirabolano Rosso (MIR) were included as outgroup and compared with the database of all *P. domestica* varieties from Sardinia to verify similarities. These two samples were well distinguished from the other with high values of classification respectively of 94.4% for MIB and 85.6% for MIR (Table 6).

3.4 The most discriminant features

In the evaluation of the parameters, in Table 5 the most discriminant five variables, over the 25 selected and used by the stepwise LDA, are reported. The first four variables are densitometric features, descriptive of the endocarp color, while the fifth is a shape descriptive variable. As a whole, 17 of the 25 chosen variables are color descriptive (data not shown).

Table 5 Ranking of the best five discriminant morpho-colorimetric variables selected and used by the LDA.

	Tolerance	F-to-remove	Wilks’ lambda
SqD_{sum}	0.002	126.714	0.096
D_{sum}	0.002	101.204	0.091
S_{mean}	0.017	55.403	0.083
E	0.119	55.290	0.083

Com	0.053	52.477	0.082

Table 6 Comparisons among the analysed plums cultivars. The percentages of correct identification are given in bold; the number of endocarps are given in parentheses.

	BOS	CAD	CAR	COR	CCO	CRO	DOA	BON	GIB	LUI	LA1	LA2	LA3	LA4	LA5	MEL	MIG	MIR	~~.3 (3)~~	NES	PAR	SAG	SB1	SB2	SAE	SIG	Total
BOS	**100.0 (72)**	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	100.0 (72)
CAD	–	**97.5 (78)**	–	–	–	–	–	–	–	–	–	–	2.5 (2)	–	–	–	–	–	–	–	–	–	–	–	–	–	100.0 (80)
CAR	–	–	**82.1 (32)**	–	–	–	–	–	5.1 (2)	–	–	–	–	–	–	–	–	–	–	–	–	–	5.1 (2)	5.1 (2)	–	2.6 (1)	100.0 (39)
COR	–	–	–	**78.8 (67)**	14.1 (12)	–	–	–	–	–	–	–	–	–	3.5 (3)	–	–	–	–	1.2 (1)	–	–	–	–	–	2.4 (2)	100.0 (85)
CCO	–	–	–	30.0 (21)	**60.0 (55)**	–	–	–	–	–	–	–	–	–	–	2.5 (2)	–	–	2.5 (2)	–	–	–	–	–	–	–	100.0 (80)
CRO	–	–	–	–	–	**90.0 (55)**	1.7 (1)	–	–	–	–	–	–	–	5.0 (5)	–	–	–	3.3 (2)	–	–	–	–	–	–	–	100.0 (60)
DOA	–	–	–	–	–	–	**96.7 (29)**	–	–	–	–	–	–	–	–	–	–	3.3 (1)	–	–	–	–	–	–	–	–	100.0 (30)
BON	–	–		–	–	–	–	**89.0 (89)**	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	10.0 (10)	100.0 (100)
GIB	–	–	–	–	–	–	–	–	**85.0 (51)**	1.7 (1)	–	–	–	–	–	–	–	1.7 (1)	–	–	–	–	5.0 (3)	5.0 (3)	–	1.7 (1)	100.0 (60)
LUI	–	–	–	–	–	3.3 (1)	–	–	–	**90.0 (27)**	–	–	–	–	–	–	–	–	–	–	–	–	3.3 (1)	–	–	3.3 (1)	100.0 (30)
LA1	–	–	–	–	5.6 (6)	–	–	–	–	–	**92.2 (83)**	1.1 (1)	–	–	1.1 (1)	–	–	–	–	–	–	–	–	–	–	–	100.0 (90)
LA2	–	–	–	–	–	–	–	3.3 (3)	–	–	–	**88.9 (80)**	–	–	1.1 (1)	–	–	–	–	–	–	–	–	4.4 (4)	–	1.1 (1)	100.0 (90)
LA3	–	–	–	–	–	–	–	–	–	–	–	–	**100.0 (85)**	–	–	–	–	–	–	–	–	–	–	–	–	–	100.0 (85)
LA4	–	–	–	–	–	–	–	–	–	3.3 (1)	–	–	–	**96.7 (29)**	–	–	–	–	–	–	–	–	–	–	–	–	100.0 (30)
LA5	–	–	–	–	–	5.7 (4)	1.4 (1)	1.4 (1)	1.4 (1)	–	–	–	–	1.4 (1)	**72.9 (51)**	–	–	–	–	1.4 (1)	–	5.7 (4)	2.9 (2)	1.4 (1)	4.3 (3)	–	100.0 (70)
MEL	–	–	–	–	–	1.7 (1)	–	–	–	–	–	–	–	–	–	**94.4 (85)**	–	–	–	–	2.2 (2)	2.2 (2)	–	–	–	–	100.0 (90)
MIG	–	–	1.1 (1)	–	–	–	3.3 (3)	–	–	–	–	–	–	–	–	–	**94.4 (85)**	1.1 (1)	–	–	–	–	–	–	–	–	100.0 (90)
MIR	–	–	4–4	–	–	–	2.2	–	1.1	–	–	–	–	–	–	–	–	**85.6**	–	–	–	–	3,3	1.1	–	2.2	100.0

			(3)				(2)		(1)									**(77)**						(1)		(2)	(90)
LIM	–	–	–	–	–	17 (1)	–	–	–	–	3.3 (2)	–	–	–	–	–	–	–	**90.0 (54)**	5.0 (3)	–	–	–	–	–	–	100.0 (60)
NES	–	–	–	–	–	–	–	1.0 (1)	1.0 (1)	–	–	–	–	–	3.0 (3)	–	–	–	–	**88.0 (88)**	–	4.0 (4)	–	–	2.0 (2)	–	100.0 (100)
PAR	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	**94.4 (17)**	55.6 (1)	–	–	–	–	100.0 (18)
SAG	–	–	–	–	–	–	–	–	–	3.5 (2)	–	–	–	1.8 (1)	–	1,8	–	–	1.8 (1)	8.8 (5)	–	**68.4 (39)**	–	–	14.0 (8)	–	100.0 (57)
SB1	1.1 (1)	–	5.6 (3)	–	–	–	3.3 (2)	–	5.6 (6)	–	–	–	–	–	3.3 (3)	–	–	1.1 (1)	–	1.1 (1)	–	–	**67.8 (61)**	7.8 (7)	–	3.3 (3)	100.0 (90)
SB2	–	–	8.3 (5)	–	–	–	1.7 (1)	1.7 (1)	6.7 (4)	–	–	–	–	–	–	–	–	1.7 (1)	–	–	–	–	8.3 (5)	**68.3 (41)**	–	3.3 (2)	100.0 (60)
SAE	1.1 (1)	–	–	–	–	–	–	–	1.1 (1)	–	–	–	–	–	7.8 (7)	–	–	–	–	1.1 (1)	–	2.2 (2)	1.1 (1)	–	**84.4 (76)**	1.1 (1)	100.0 (90)
SIG	–	–	1.1 (1)	1.1 (1)	–	–	1.1 (1)	2.2 (2)	–	–	–	–	–	–	1.1 (1)	–	–	–	–	–	–	–	–	2.2 (2)	–	**91.1 (82)**	100.0 (90)
Overall																											**86.1% (1836)**

4 Discussion

The morpho-colorimetric analysis applied on *Prunus* endocarps proves the effectiveness of the image analysis. The macro, *Prunus*.mcr, build for endocarp analysis was able to correctly discriminate *P. domestica* varieties from wild populations of *P. spinosa* with a high degree of precision.

Sardinian fruit heritage is characterized by a huge number of cultivated varieties, but due to the great historical and anthropological heterogeneity of the island, many of these are the product of linguistic distortion creating a wide assortment of names (et al., 2016). This is what happens also for plum cultivars.

Analysing the endocarps collected for each variety, the results obtained have allowed to hypothesize the existence of two synonymy groups: The first group includes Coru (COR) and Coru e Columbu (CCO) and the second one Sanguigna di Bosa 1 (SB1) and Sanguigna di Bosa 2 (SB2).

The varieties of the first group were found in the territory of Laconi (Central Sardinia) and spread in different locations in Sardinia: they are traditionally consumed both fresh and dried.

In addition, considering the percentages of correct identification of the varieties Sanguigna di Bosa 1 (SB1) and Sanguigna di Bosa 2 (SB2) in the global comparison, it is possible to assume that, although some differences exist from the phenotypic point of view, such as flowering time, leaves, bearing shaft, disease resistance, differences in the chemical composition of anthocyanins and differences in ripening, they have a parental line in common or may be the same variety.

The results about the comparison between the commercial samples Mirabolano Giallo (MIB) and Mirabolano Rosso (MIR), with the varieties of *P. domestica* of Sardinia suggest that these varieties differ significantly and may be considered as autochthonous.

Considering the differences among cultivated varieties, the results obtained from the analysis confirm the extreme phenotypical, and more extensively biological diversity of *P. domestica.*

Regarding the discriminating power of the features is important to underline that the Elliptical Fourier shape Descriptors (EFDs) were not included by the classifier in the main parameters of discrimination, but among the parameters that influenced the discrimination process of the studied plums, the most important chosen by the stepwise LDA were mainly parameters descriptive of color and densitometry.

According to Lo Bianco et al. (2015a), the Haralick's parameters resulted to be among the most discriminant, although different species can show different discriminant characters.

Finally, based on all the results obtained, this study confirms that according to Depypere et al. (2007) the endocarp of *P. domestica* is the most stable of the all characters used for their identification at specific level.

5 Conclusions

The ancient fruits widespread in Sardinia are well adapted to different soil and climate environments representing traditionally important genetic resource destined to disappear due to anthropization.

In this sense, the field catalogues represent valid strategies to pursue sustainable quality and typicality, and contemporarily counteract the negative environmental impact of the varieties that generally are not valid from the agronomic point of view due to their different qualitative and quantitative aspects of production.

In pomology and agrobotany, plum varieties are usually characterized by their flower (e.g. diameter, color of the anthers) and fruit characteristics (e.g. color, size, shape, endocarp shape, endocarp adhesion). The morphometric study presented in this work constitutes an innovative contribution to characterize plum agrobiodiversity in Sardinia. In fact, for the first time, it was possible to investigate about the morphology and morphometry of plum endocarps of all traditional varieties from Sardinia. The 134 morpho-colorimetric and texture features measured on the germplasm resulted a valid tool to achieve a clear discrimination among different varieties, also allowing the identification of possible synonymy groups. The obtained results prove that image analysis techniques can be considered as a useful tool in taxonomic investigations, also at varietal level. This system results a good method for the quick and cheap identification of traditional plums for consumer satisfaction and for those varieties of plums that have a particular economic value, this detective procedure could be used to define objective parameters useful in the attribution of European trademarks such as "Denominazione di Origine Protetta" (DOP) and "Indicazione Geografica Protetta (IGP). This procedure could be helpful in germplasm banks, nurseries or in those institutions that deal with ex situ conservation of plant biodiversity.

In the future, it would be interesting to assess whether these results would be confirmed by molecular analysis on the same traditional varieties.

Acknowledgments

I wish to thank Mario Schirra, Giuseppe Rosas and Liberato Fara of the CNR-ISPA (Nuraxinieddu, OR-Sardinia) for scientific and technical assistance during the samples collection.

References

Agabbio, M., D'hakkewin, G., Molinu, M.G., Bacchetta, G., 2016. Educazione alla biodiversità. Antiche varietà di frutta della Sardegna. Il susino. ISPA-LAORE, 1.

Athanasiadis I., Nikoloudakis N. and Hagidimitrion M., Genetic relatedness among cultivars of the Greek Plum Germplasm, *Not. Bot. Horti. Agrobo.* **41**, 2013, 491–498.

Atzeri P., Bacchetta G., Casti M., Fenu G., Mattana E., Meloni F., Mulè P., Orrù M., Pontecorvo C. and Sarigu R., La Banca del Germoplasma della Sardegna (BG-SAR), *Studi Trentini di Scienze Naturali Acta Biologica* **88** (2012), 2012, 69–75.

Bacchetta G., Grillo O., Mattana E. and Venora G., Morpho-colorimetric characterization by image analysis to identify diaspores of wild plant species, *Flora* **203**, 2008, 669–682.

Bouby L., Figueiral I., Bouchette A., Rovira N., Ivorra S., Lacombe T., Pastor T., Pico S., Marinval J. and Terral J.F., Bioarchaeological insights into the process of domestication of grapevine (*Vitis vinifera* L.) during Roman times in Southern France, *PLoS ONE* **8**, 2013, e63195.

Depypere L., Charele P., Vander, Mijnsbrugge K. and Goetghbeur P., Stony endocarp dimension and shape variation in *Prunus* section *Prunus*, *Ann. Bot.* **100**, 2007, 1585–1597.

Faust M. and Suranyi D., Origin and dissemination of plum, *Hort. Rev.* **23**, 1999, 179–231.

Fawzi N.M., Macro and micromorphological seed characteristics of some selected species of Caesalpinioideae-Leguminosae, *Res. J. Bot.* **6**, 2011, 68–77.

Feemster Jashemski W.M. and Meyer F.G., The Natural History of Pompeii, 2002, Cambridge University Press; Cambridge.

Fisher R.A., The precision of discriminant functions, *Ann. Eugen.* **10**, 1940, 422–429.

Fisher R.A., The use of multiple measurements in taxonomic problems, *Ann. Eugen.* **7**, 1936, 179–188.

Gresta F., Santonoceto C., Ceravolo G., Formantici C., Grillo O., Ravalli C. and Venora G., Productive, qualitative and seed image analysis traits of guar (Cyamopsis tetragonoloba (L) Taub), *Aust. J. Corp. Sci.* **10** (7), 2016, 1052–1060.

Grillo O., Draper D., Venora G. and Martínez-Laborde J.B., Seed image analysis and taxonomy of *Diplotaxis* DC. (Brassicaceae, Brassiceae), *Syst. Biodivers.* **10**, 2012, 57–70.

Grillo O., Mattana E., Fenu G., Venora G. and Bacchetta G., Geographic isolation affects inter- and intra-specific seed variability in the *Astragalus tragacantha* complex, as assessed by morpho-colorimetric analysis, *C. R Biol.* **336**, 2013, 102–108.

Grillo O., Mattana E., Venora G. and Bacchetta G., Statistical seed classifiers of 10 plant families representative of the Mediterranean vascular flora, *Seed Sci. Technol.* **38**, 2010, 455–476.

Grillo O., Miceli C. and Venora G., Computerised image analysis applied to inspection of vetch seeds for varietal identification, *Seed Sci. Technol.* **39**, 2011, 490–500.

Haralick, R.M., Shanmugam, K., Its'hak, D., 1973. Textural features for image classification. IEEE. S.M.C. 3, pp. 610–621.

Haralick R.M. and Shapiro L.G., Glossary of computer vision terms, *Pattern Recogn.* **24**, 1991, 69–93.

Hâruta O., Elliptic Fourier analysis of crown shapes in Quercus petraea trees, *Ann. For. Res.* **54**, 2011, 99–117.

Hastie T., Tibshirani R. and Friedman J., The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2001, Springer; New York.

Holden J.E., Finch W.H. and Kelley K., A comparison of two-group classification methods, *Educ. Psychol. Meas.* **71**, 2011, 870–901.

Horvath A., Balsemin E., Barbot J.C., Christmann H., Manzano G., Reynet P., Laigret F. and Mariette S., Phenotypic variability and genetic structure in plum (*Prunus domestica* L.), cherry plum (*P. cerasifera* Ehrh.) and sloe (*P. spinosa* L.), *Sci. Hortic.* **129**, 2011, 283–293.

Janick J., The origin of fruits, fruit growing and fruit breeding, *Plant Breed. Rev.* **25**, 2005, 255–320.

Kuhn M. and Johnson K., Discriminant analysis and other linear classification models, In: *Applied Predictive Modeling,* 2013, Springer; New York, 275–328.

Lo Bianco, M., Ferrer-Gallego, P., Grillo, O., Laguna, E., Venora, G., Bacchetta, G., 2015a. Seed image analysis provides evidence of taxonomical differentiation within the Medicago L. sect. Dendrotelis (Fabaceae). Syst. Biodivers. DOI: http://dx.doi.org/10.1080/14772000.2015.1046968.

Lo Bianco M., Grillo O., Cremonini R., Sarigu M. and Venora G., Characterization of Italian bean landraces (*Phaseolus vulgaris* L.) using seed image analysis and texture descriptors, *Aust. J. Crop Sci.* **9**, 2015b, 1022–1034.

Mattana E., Grillo O., Venora G. and Bacchetta G., Germplasm image analysis of *Astragalus maritimus* and *A. verrucosus* of Sardinia (subgen. Trimeniaeus, Fabaceae), *Ann. Jard. Bot. Madrid* **65**, 2008, 149–155.

Nielsen J. and Olrik D.C., A morphometric analysis of *Prunus spinosa*, *P. domestica* ssp. *insititia*, and their putative hybrids in Denmark, *Nord. J. Bot.* **21**, 2001, 349–363.

Orrù M., Grillo O., Venora G. and Bacchetta G., Computer vision as a method complementary to molecular analysis: Grapevine cultivar seeds case study, *C. R. Biol.* **335**, 2012, 602–615.

Orrù M., Grillo O., Venora G. and Bacchetta G., Seed morpho-colorimetric analysis by computer vision: a helpful tool to identify grapevine (Vitis vinifera L.) cultivars, *Aust. J. Grape Wine Res.* **21**, 2015, 508–519.

Pinna M.S., Grillo O., Mattana E., Cañadas E.M. and Bacchetta G., Inter- and intraspecific morphometric variability in *Juniperus* L. seeds (Cupressaceae), *Syst. Biodivers.* **12**, 2014, 211–223.

Piras F., Grillo O., Venora G., Lovicu G., Campus M. and Bacchetta G., Effectiveness of a computer vision technique in the characterization of wild and farmed olives, *Comput. Electron. Agric.* **122**, 2016, 86–93.

Pollmann B., Jacomet S. and Schlumbaum A., Morphological and genetic studies of waterlogged *Prunus* species from the Roman *vicus Tasgetium* (Eschenz, Switzerland), *J. Archaeol. Sci.* **32**, 2005, 1471–1480.

Rovner I. and Gyulai F., Computer-assisted morphometry: a new method for assessing and distinguishing morphological variation in wild and domestic seed populations, *Econ. Bot.* **61**, 2007, 154–172.

Sabato D., Esteras C., Grillo O., Picó B. and Bacchetta G., Seeds morpho-colorimetric analysis as complementary method to molecular characterization of melon diversity, *Sci. Hortic.* **192**, 2015, 441–452.

Santo A., Mattana E., Grillo O. and Bacchetta G., Morpho-colorimetric analysis and seed germination of *Brassica insularis* Moris (Brassicaeae) populations, *Plant Biol.* **17**, 2015, 335–343.

Shahin M.A. and Symons S.J., Colour calibration of scanners for scanner independent grain grading, *Cereal Chem.* **80**, 2003, 285–289.

Smykalova I., Grillo O., Bjelkova M., Hybl M. and Venora G., Morpho-colorimetric traits of *Pisum* seeds measured by an image analysis system, *Seed Sci. Technol.* **39**, 2011, 612–626.

Smykalova I., Grillo O., Bjelkova M., Pavelek M. and Venora G., Phenotypic evaluation of flax seeds by image analysis, *Ind. Crop Prod.* **47**, 2013, 232–238.

SPSS Base 16.0, Application Guide, 2006, Prentice Hall; New Jersey.

Sugiyama M., Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis, *J. Mach. Learn. Res.* **8**, 2007, 1027–1061.

Ucchesu M., Orrù M., Grillo O., Venora G., Paglietti G., Ardu A. and Bacchetta G., Predictive method for correct identification of archaeological charred grape seeds: support for advances in knowledge of grape domestication process, *PLoS ONE* **11** (2), 2016, e0149814.

Ucchesu M., Orrù M., Grillo O., Venora G., Usai A. and Serreli P.F., Earliest evidence of a primitive cultivar of *Vitis vinifera* L. during the Bronze age in Sardinia (Italy), *Veg. Hist. Archaeobot.* **24**, 2015, 587–600.

Usenik V., Štampar F. and Veberič R., Anthocyanins and fruit colour in plums (*Prunus domestica* L.) during ripening, *Food Chem.* **114**, 2009, 529–534.

Zohary, D., Hopf, M., Weiss, E., 2012. Domestication of plants in the old world. The origin and spread of cultivated plants in West Asia, Europe and the Nile valley. Oxford University Press, Oxford.

Highlights

- *P. domestica* endocarps biometric features were measured by image analysis techniques.
- A total of 134 size, shape, color and texture descriptors were measured on each endocarp.
- The data were statistically processed applying the stepwise Linear Discriminant Analysis.
- Hypothetical synonymy groups were identified.
- This is the first computer vision application on plum biodiversity study.

Queries and Answers

Query: Please check the edit and layout of Table 6, and correct if necessary.
Answer: Please change ".3 (3)" with "LIM" in the first line of the table.

Query: Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact s.sekar@elsevier.com immediately prior to returning your corrections.
Answer: The article is correctly registered as a regular item.

Query: The author names have been tagged as given names and surnames (surnames are highlighted in teal color). Please confirm if they have been identified correctly.
Answer: Correct

Query: Please check edits made in affiliations ’a-e’, and correct if necessary.
Answer: Correct

Query: Please check the address of the corresponding author, and correct if necessary.
Answer: Correct

Query: Reference ‘Eryomine (1990)’ is cited in the text but not provided in the reference list. Please provide it in the reference list or delete these citations from the text.
Answer: It can be deleted

Query: The citations ‘Faust and Surányi (1999)’ and Feemster and Meyer (2002)’ have been changed to match in the reference list. Please check here and in subsequent occurrences, and correct if necessary.
Answer: Correct

Query: Please provide a caption for Fig. 1.
Answer: Fig. 1 Graphic representation of some morphometric measurements on a *Prunus domestica* seed.

Query: Please check the clarify the cross reference citation ‘et al. (2016)’ in sentence ‘Sardinian fruit heritage is…’, and correct if necessary.
Answer: It is Agabbio et al., 2016