Computers and Electronics in Agriculture 122 (2016) 86–93

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

Original papers

Effectiveness of a computer vision technique in the characterization of wild and farmed olives

Fabio Piras [a], Oscar Grillo [b,*], Gianfranco Venora [c], Gianni Lovicu [a], Marco Campus [a], Gianluigi Bacchetta [b]

[a] Agris Sardegna – Agenzia per la ricerca in agricoltura, Loc. Bonassai S.S. 291 Sassari-Fertilia, km. 18,600, Sassari, Italy
[b] Centro Conservazione Biodiversità, Dipartimento di Scienze della vita e dell'ambiente, Università di Cagliari, V.le S. Ignazio da Laconi 13, 09123 Cagliari, Italy
[c] Stazione Consorziale Sperimentale di Granicoltura per la Sicilia, Via Sirio 1, 95041 Santo Pietro, Caltagirone (CT), Italy

ARTICLE INFO

Article history:
Received 20 October 2015
Received in revised form 8 December 2015
Accepted 19 January 2016
Available online 3 February 2016

Keywords:
Image analysis
Landraces
Morpho-colorimetric characterization
Olea europaea L.
Sardinia
Wild crop relatives

ABSTRACT

Existing in two forms, namely the wild and the cultivated one, the olive tree (*Olea europaea* L.) is one of the most widespread species in the Mediterranean Basin since the sixth millennium B.C. According to the theory of a polycentric origin for domestication, mainly supported by the presence of the wild form all over the Mediterranean area, the goal of this study is to use endocarps biometric features, measured by image analysis techniques, in order to investigate the relationships among different Sardinian wild olive populations, ancient olive trees and olive cultivars. A total of 137 morphometric, colorimetric and textural characters were measured on each of the 2842 studied endocarps. The morpho-colorimetric analysis of olive endocarps proved the effectiveness of the technique to discriminate between the two currently recognized *O. europaea* L. varieties (var. *europaea* and var. *sylvestris*). The partial overlapping that resulted between wild and ancient olives suggests the existence of large populations characterized by a high degree of phenotypic variability. Similar results were reached for varietal discrimination, perfectly fitting with previous findings of molecular studies. Considering the phenotypic and genetic distance of cultivars from autochthonous wild olives, the allochthonous origin of local varieties is furthermore confirmed.

1. Introduction

The olive tree (*Olea europaea* L.), a widespread species in the Mediterranean Basin, exists in two forms, namely the wild [var. *sylvestris* (Mill.) Lehr] and the cultivated one (var. *europaea*) (http://www.theplantlist.org). The former represents one of the main structural species of the Mediterranean scrublands (Bacchetta et al., 2003), while the latter is the most widely cultivated tree in the area, holding an important historical and cultural relevance (Zohary et al., 2012).

The existence of wild olive trees can be traced back to Paleolithic in Near East, according to palynological and antracological studies (Liphschitz et al., 1991), while during the Neolithic age, man began to use wild olive fruit in his diet (Liphschitz et al., 1991; Zohary and Hopf, 1993). Archaeological evidence for intensive exploitation of olive fruit is already present from the sixth millennium B.C., in Near East sites (Zohary and Spiegel-Roy, 1975; Liphschitz et al., 1991).

Moreover, evidence based on genetic studies supports the theory of a polycentric origin for domestication (Claros et al., 2000; Bronzini de Caraffa et al., 2002a; Rotondi et al., 2003; Besnard et al., 2001, 2013). This theory is also supported by the high number of varieties existing in the countries where the olive is historically exploited (Rallo, 2005), as well as by the presence of the wild form all over the Mediterranean Basin.

From the cradle of first domestication in the Northern Levant, the dispersal across the Mediterranean Basin occurred in parallel with the expansion of civilizations and human exchanges in this part of the world. Dispersal from original sites of domestication by migration and human exchanges is supported by the genetic distance between varieties and local populations of the wild tree (Angiolillo et al., 1999; Baldoni et al., 2006).

Sardinia, the second largest island of the Mediterranean Basin, represents a peculiar field of investigation on this topic. Indeed, numerous ancient olive trees, both wild and cultivated, grow in the region, including the oldest Italian ancient tree called "olivastro di Luras" (Vannelli, 1989). Ancient specimens represent exceptionally interesting materials for the study of evolutionary processes, domestication and diffusion of the olive tree (Baldoni et al., 2006). Moreover, the island has undergone important historical events. As a part of the Spanish Empire for about three centuries, Sardinia has greatly promoted the introduction of cultivars and the cultivation of olive, and its geographical position represents

* Corresponding author.
E-mail address: oscar.grillo.mail@gmail.com (O. Grillo).

http://dx.doi.org/10.1016/j.compag.2016.01.021
0168-1699/

the cut-off between the Eastern and Western Mediterranean oleaster lineages (Besnard et al., 2002).

In the last two decades, a significant increase in image analysis applications has been highlighted in plant biology research to quantify the phenotypic diversity in wild and cultivated vascular plant species (Kiliç et al., 2007; Bacchetta et al., 2011a,b; Appelhans et al., 2011; Fawzi, 2011; Herridge et al., 2011; Grillo et al., 2011, 2013; Smykalova et al., 2011, 2013; Pinna et al., 2014; Lo Bianco et al., 2015a; Santo et al., 2015) and to provide a starting point to identify well-preserved archaeological seed materials (Sabato et al., 2015a; Orrù et al., 2013; Terral et al., 2010; Ucchesu et al., 2014). In a recent paper, Sabato et al. (2015b) analyzed several accessions of melon, belonging to different genera and species. They compared the data from seed phenotypic characterization, achieved by image analysis techniques, with those from molecular analysis on melon genotypes, in order to highlight seed morphology distances among genus and species and to prove, yet again, the high discriminant power of seeds.

Following the same morpho-colorimetric approach, the aims of the present study are to:

- build a database of phenotypic features of wild and farmed olive endocarps;
- implement statistical classifiers able to compare and discriminate between cultivated varieties and wild populations;
- investigate the relationships among different wild olive populations, ancient olive trees and olive cultivars.

This study represents the first attempt at morpho-colorimetric characterization applied to *O. europaea* L., based on endocarps image analysis.

2. Materials and methods

2.1. Sample lots details

Fruits from wild olives were collected from 16 populations, selected for their maximum distance and isolation from cultivated areas, in order to avoid the sampling of hybrid material (Table 1) as much as possible. From each population, 12 fruits were randomly picked from five trees (an overall of 60 drupes were collected for each population), from all over the canopy. Applying the same sampling scheme, a further 22 cultivars were sampled and analyzed (Table 1). Part of the material was collected directly in the field and other samples were collected from olive orchards located in the main growing areas, and the remaining samples were collected from the olive varieties collections of Agris Sardegna (Agricultural *Research* Agency of *Sardinia*).

Samples constituting 12 fruits were collected from ancient olive trees, sampling 47 accessions in 15 different locations (Table 1). The trees were chosen according to their dimensions (from 3 to 12.6 m of circumference measured at 1.2 m from the ground), being presumptively genuine wild olives of old age. Sampling areas for wild, ancient trees and cultivars are reported in Fig. 1.

All the samples were collected in autumn and winter, with fruits at full ripeness, according to exocarp pigmentation (dark purple), and endocarps at the complete morphologic and chromatic development. Exocarps and mesocarps were manually removed, and the endocarps accurately cleaned, washed and dried.

2.2. Endocarps image analysis

Digital images of endocarps samples were acquired using a flatbed scanner (Epson V600 Perfection) with a digital resolution of 400 dpi and a scanning area not exceeding 1024 × 1024 pixels. Before image acquisition the scanner was calibrated for color matching, following the protocol of Shahin and Symons (2003), as suggested by Venora et al. (2007a). Digital images of samples consisting of 12 endocarps, randomly disposed on the flatbed tray, were acquired and used for the analysis. Each sample was acquired two times, picking up and re-arranging the endocarps on the scanner tray.

The captured images were processed and analyzed using the software package KS-400 V.3.0 (Carl Zeiss, Vision, Oberkochen, Germany). This software can be customized for specific applications by editing appropriate image analysis algorithms in "macros", allowing to automatically accomplish all the analysis procedures, reducing the execution time and contextual mistakes in the analysis process. The accuracy and speed of measurements was maximized by running an automated macro specifically developed for the characterization of wild endocarps (Bacchetta et al., 2008a; Mattana et al., 2008; Grillo et al., 2010). In order to increase the number of discriminant parameters, the Elliptic Fourier Descriptors (EFDs) were also computed as described by Orrù et al. (2012). Moreover, following the same procedure described by Lo Bianco et al. (2015b), the macro was further improved by adding a series of algorithms able to compute 11 Haralick's descriptors (Haralick et al., 1973) and the relative standard deviations for each analyzed endocarp. The 11 Haralick's descriptors measured on each seed to mathematically describe the surface texture, are reported as Supplementary Material (Suppl. Table 1).

A total of 137 morphometric, colorimetric and textural characters were measured on each endocarp (Suppl. Table 2; Suppl. Fig. 1).

2.3. Statistical analysis

The achieved results were used to build a database including morpho-colorimetric, EFDs and Haralick's descriptors. Statistical elaborations were executed using SPSS software package release 16.0 (SPSS Inc. for Windows, Chicago, Illinois, USA), applying the same stepwise linear discriminant analysis (LDA) algorithm suggested by Grillo et al. (2012) to identify and discriminate among the investigated endocarps accessions. This approach is commonly used to classify/identify unknown groups characterized by quantitative and qualitative variables (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; Venora et al., 2009; Holden et al., 2011; Rencher and Christensen, 2012; Kuhn and Johnson, 2013). The original LDA formulation, also known as the Fisher Linear Discriminant Analysis (FLDA) (Fisher, 1936, 1940) deals with binary classifications. The selection of the original features is carried out by a stepwise procedure able to identify and select the most statistically significant features among them to use for the endocarps samples identification. The process is automatically stopped when no remaining measured features increased the discrimination ability (Venora et al., 2007b; Grillo et al., 2012).

The leave-one-out cross-validation (LOOCV) procedure was applied to verify the performance of the identification system, testing individual unknown cases and classifying them on the basis of all others. It involves using a single case from the original sample set as the validation dataset, and the remaining cases as the training set. Each case is classified into a group according to the classification functions computed from all the data except the case being classified. The proportion of misclassified cases, after removing the effect of each case one at a time, is the leave-one-out estimate of misclassification (SPSS, 2007).