Computers and Electronics in Agriculture 141 (2017) 223–231

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

Original papers

Wheat landraces identification through glumes image analysis

Oscar Grillo *, Sebastiano Blangiforti, Gianfranco Venora

Stazione Consorziale Sperimentale di Granicoltura per la Sicilia, Via Sirio 1, Borgo Santo Pietro, Caltagirone (CT), Italy

A R T I C L E I N F O

Article history:
Received 6 June 2017
Received in revised form 21 July 2017
Accepted 27 July 2017

Keywords:
Biodiversity
Local populations
Morpho-colorimetric analysis
Old varieties
Triticum L

A B S T R A C T

A new practical method able to identify wheat local landraces was implemented. It is based on computerized image analysis techniques and statistical identification, for the first time on the basis of glumes size, shape, colour and texture.

Ears of 52 different Sicilian wheat landraces were reaped for three consecutive years. Digital images of the glumes were acquired, processed and analysed, measuring 138 quantitative morpho-colorimetic variables. The data were statistically analysed applying a Linear Discriminant Analysis. All the statistical comparisons, distinguished for systematic rank, given perfect identification performances; while an overall percentage of correct identification of 89.7% was reached when all the landraces were compared all together.

Finally, the identification system was tested with an unknown glume sample, later entirely identified as Vallelunga, one of the Sicilian landraces.

This work represents the first attempt of wheat landraces identification based on glume phenotypic characters, applying image analysis techniques. Considering the growing interest in local old wheat landraces, strongly linked to the renewed appreciation in traditional and typical local products, the obtained results support the application of the image analysis system not only for grading purposes, but also to define the product traceability, in order to get a "market card" for wheat landraces.

1. Introduction

Wheat (*Triticum* subsp.) is one of the main food sources in the world. Its world production for 2016/17 is approximately expected in 740 million tons, exceeding the 2015/16 record by 1.2%, and covering about 15% of the world's arable surface (FAO, 2017). Durum wheat production reaches around 30 million tons in about 16 million hectares, accounting approximately 5–6% of the total world wheat production (Cebola Lidon et al., 2014). It is commonly grown in most of the countries around the world, although the Mediterranean region produces about 60% of world durum wheat production (Morancho, 2000), being the EU (Italy, Spain, France and Greece) the leading global producer (Cebola Lidon et al., 2014). On this scenario, south Italy is one of the regions historically most voted to the cereal crops, where the durum wheat varietal biodiversity is particularly high.

Sicily, with an area of 25,711 km^2, is the largest island in the Mediterranean sea and due to its geographical position and extremely diversified ecological condition, always hosted an ideal environment for the cultivation of cereals and in particular durum wheat. This is due to the extreme variability of altitude and pedo-climatic conditions, characterized by clayish to sandy fields, by variable orography, distance from sea and wind regime (Lombardo, 2004). Some socio-cultural aspects had also contributed enriching the varietal heritage, such as the great amount of invasions that, during the centuries, conquered wide Sicilian areas, favoured by the strategic geographical position of the island. All these conditions, together with the mass selection historically conducted and the more recent genetic improvement programs based on artificial crosses, had contributed to build the extremely wide varietal panorama currently existing. On the other hand, Sicily is known as "Republic granary" since III-II century b.C., as reported by Caton the censor (234–149 b.C.).

In Sicily are currently cropped a few of tens of old and new durum wheat varieties officially recorded and regulated with national and communitarian protocols, but also many ancient landraces or populations characterized by specific bio-morphological traits and qualitative features (Spina et al., 2008; Sciacca et al., 2014). A cropped variety or cultivar, is an intra-specific taxonomic entity characterized by high level of homozygosis, specially for the genes that control the selected traits, consequently, the individuals belonging to the same variety show homogeneous morphological and/or productive traits. Nevertheless, some differences in

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

http://dx.doi.org/10.1016/j.compag.2017.07.024

genetically controlled biochemical traits may exist within a same variety (e.g. protein components) (Peruffo et al., 1985). These variations were defined "biotypes". Differently from varieties, landraces are natural populations put in cultivation and as such, they are characterize by wide adaptability to various environments including irrigated and dry land conditions (Jones et al., 2008; Camacho et al., 2005). Considering all the abiotic factors, the high probability of inter-population crosses and their heterozygosis condition, from the genetic point of view, these populations result to be more than a mixture of different pure lines (Zeven, 1998; Landjeva et al., 2015).

Up to 100–150 years ago, the landraces were the only one kind of wheat cultivar available for the farmers; afterward, knowledge and new technologies launched the genetic improvement as understood today. It reflected on a marked genetic and phenotypic homogeneity, useful for the mechanization of many agronomic practices; but it also reflected on a greater phenological synchrony, helpful for the application of herbicides, pesticides and fertilizers. Homogeneity also positively affects on the value of the productions addressed to industry, and from the legal point of view, facilitates the unequivocal varietal identification.

The extreme homogeneity also implicates the negative aspects related to the biotic and abiotic stresses. Wild plants are indeed unlikely subjected to epidemics and pathogenic attacks. Moreover, changing old varieties genetically heterogeneous with new ones certainly more homogeneous, damaging local populations, activates genetic erosion phenomena (Guarda et al., 2004; Newton et al., 2010).

Broadly speaking, genetic improvement has always existed, but what has changed in the last century is not only the nature of the selection, but the nature and range of genetic variability (Frankel, 1970).

At the beginning of the XX century, in Sicily as well as in the rest of the world, the strong interest in biodiversity conservation powered up the accurate search of cropped species germplasm. In Europe, many researchers started to collect ex-situ seed materials in germplasm banks. Vavilov (1957), at time one of the most operative investigators, found many seed material belonging to cultivated plant species, establishing the origin and speciation centres of a great part of the currently cropped species. Thanks to the work of these scientists, a lot of endangered local varieties were saved and currently made available for breeding programs and typical products making.

In recent years, in Sicily as well as in the rest of Europe, the attention paid to local and traditional productions and is growing, especially in the agro-food sector. For economic, social and nutritional reasons, this trend has led to the rediscovery and reuse of landraces both of wheat and other crops, responding to requests for more and more demanding market. The rising price of these local productions and the consequent increased satisfaction of farmers, is proving to be an interesting professional opportunities also for young workers. Moreover, many recent studies testify the high healthy and nutraceutical value of old landraces, both for high amount of antioxidant compounds and for their natural aptitude to organic production (Gallo et al., 2004; Pasqualone et al., 2014; Migliorini et al., 2016; Lo Bianco et al., submitted for publication).

This growing interest in local old landraces has inspired to find effective and objective identification methods, able to distinguish old landraces (Grillo et al., 2016).

In the recent past, many DNA-based methods have been set up, for wheat-derived products, to trace cultivars in starting seed stocks, semolina, bread and pasta (Pasqualone et al., 1999, 2000; Fujita et al., 2009). Giancaspro et al. (2016) described the denaturing high performance liquid chromatography technique for setting up a single nucleotide polymorphism based method to achieve the varietal traceability of the durum wheat cultivar "Timilia", reaching no very high but promising percentage of detection.

Anyway genetic approach is not the only one. Substantial work dealing with the use of different morphological (size and shape) features for classification of wheat grains and varieties has been reported in the literature (Keefe and Draper, 1986; Zayas et al., 1989; Barker et al., 1992; Arefi et al., 2011; Zapotoczny, 2011). Modern phenotyping methods proved to be a helpful tool both in plant identification and classification and in quality assessment (Venora et al. 2009; Guevara-Hernandez and Gomez-Gil, 2011; Smykalova et al., 2011, 2013). Pourreza et al. (2011) applied machine vision techniques to classify nine common wheat varieties based on seeds; while recently, Szczypiński et al. (2015) implemented an identification system to discriminate among 11 barley varieties based on image-derived shape, colour and texture attributes of individual kernels, reaching an accuracy included between 67% and 86%. Many other researches, based on image analysis technology, were recently conducted in order to distinguish wheat and other cereal varieties (Szczypiński and Zapotoczny, 2012; Mebatsion et al., 2013; Chaugule and Mali, 2016). Although seeds and kernels proved to be the right matrix to study in order to discriminate among varieties, problems arise increasing the varietal sample amount and above all when no genetically defined samples, such as populations or landraces, have to be identified.

The aim of this paper is to establish a practical method based on computerized image analysis techniques and statistical identification capable to identify wheat local landraces, for the first time on the basis of glumes size, shape, colour and texture.

2. Material & methods

2.1. Samples details

Ears of 52 different wheat local varieties or landraces were reaped, at the time of maximum ripening, from the fields of the Stazione Sperimentale di Granicoltura per la Sicilia, sited in Santo Pietro – Caltagirone [37 °07′12″N; 14 °31′17″E; 313 m a.s.l.] (CT, Sicily, Italy) (Table 1; Fig. 1). In order to include a widest morphological and environmental variability, the wheat ears were collected during three consecutive years (2012, 2013, 2014).

From three to six ears were sampled and from two to four glumes were removed from the spikelets of the ear middle section and from the both sides of each ear. The glumes were stored at room temperature under controlled conditions (20 °C and 50% RH).

Applying the same sampling approach, one more unknown landrace, collected in 2015 from Gangi (PA, Sicily, Italy) in the Madonie mountains (C-N Sicily), locally named "Nivuru", was used to test and validate the identification system.

2.2. Glume image analysis

Digital images of glumes samples were acquired using a flatbed scanner (ScanMaker 9800 XL, Microtek Denver, CO) with a digital resolution of 400 dpi and a scanning area not exceeding 1024 × 1024 pixel. Before image acquisition, the scanner was calibrated for colour matching following the protocol of Shahin and Symons (2003) as suggested by Venora et al. (2009). Images consisting of few wheat glumes were captured, disposing them on the flatbed tray, distinguishing in right and left side of the ear and used for the digital image analysis. Morpho-colorimetric features were only measured for sound intact glumes, rejecting that ones with broken beak or shoulder. A total of 4253 wheat glumes were analysed.

Table 1
List of the 52 different wheat local varieties studied.

Code	Variety/Landrace	Species	Sample amount
bb2	Bufala Bianca 02	*T. turgidum* L.	45
bd3	Bidì 03	*T. durum* Desf.	31
bia1	Biancuccia 01	*T. durum* Desf.	40
bivc	Casedda (Bivona)	*T. turgidum* L.	24
bnc2	Bufala Nera Corta 02	*T. turgidum* L.	28
bnl1	Bufala Nera Lunga 01	*T. turgidum* L.	35
brc-b1	Bufala Rossa Corta b01	*T. turgidum* L.	40
brl1	Bufala Rossa Lunga 01	*T. turgidum* L.	92
cat	Capeiti	*T. durum* Desf.	28
cal	Cappelli	*T. durum* Desf.	12
cas1pu	Castiglione Pubescente 01	*T. durum* Desf.	98
cas3gl	Castiglione Glabro 03	*T. durum* Desf.	49
chi1	Chiattulidda 01	*T. durum* Desf.	30
cic1	Ciciredda 01	*T. turgidum* L.	31
cot1	Cotrone 01	*T. durum* Desf.	90
cuc1	Cuccitta 01	*T. aestivum* L.	99
fce1	Francesone 01	*T. durum* Desf.	98
fl3	Farro Lungo 03	*T. durum* Desf.	95
fsa1	Francesa 01	*T. durum* Desf.	97
gig1	Gigante 01	*T. durum* Desf.	95
gio1	Gioia 01	*T. durum* Desf.	95
gir1	Girgentana 01	*T. durum* Desf.	90
giu1	Giustalisa 01	*T. durum* Desf.	86
ing2	Inglesa 02	*T. durum* Desf.	95
lin1	Lina 01	*T. durum* Desf.	95
mai1pol	Maiorca di Pollina 01	*T. aestivum* L.	84
mai6	Maiorca 06	*T. aestivum* L.	76
mar2	Margherito 02	*T. durum* Desf.	90
mar6	Margherito 06	*T. durum* Desf.	97
mce2	Maiorcone 02	*T. aestivum* L.	76
m1a1	Martinella 01	*T. durum* Desf.	95
mm1	Manto di Maria 01	*T. durum* Desf.	95
pao2	Paola 02	*T. turgidum* L.	93
pav3	Pavone 03	*T. durum* Desf.	98
rea4	Realforte 04	*T. durum* Desf.	95
reg1	Regina 01	*T. durum* Desf.	93
rom2	Romano 02	*T. aestivum* L.	89
rsc9	Ruscia 09	*T. durum* Desf.	97
rus1	Russello 01	*T. durum* Desf.	81
russg8	Russello 13 SG8	*T. durum* Desf.	97
sca1	Scavuzza 01	*T. durum* Desf.	95
sco4	Scorsonera 04	*T. durum* Desf.	98
sem1	Semenzella 01	*T. durum* Desf.	98
sam3	Sammartinara 03	*T. durum* Desf.	144
sic1	Sicilia 01	*T. durum* Desf.	98
tim1	Timilia 01	*T. durum* Desf.	98
tre2	Trentino 02	*T. durum* Desf.	119
tri2	Tripolino 02	*T. durum* Desf.	80
tumsg3	Tumminia SG3	*T. durum* Desf.	94
tun1	Tunisina 01	*T. durum* Desf.	76
urr1	Urrìa 01	*T. durum* Desf.	88
val	Vallelunga	*T. durum* Desf.	191
UGS	Unknown glume sample		54

All the images were processed and analysed using the software package KS-400 V. 3.0 (Carl Zeiss, Vision, Oberkochen, Germany). A macro, specifically developed for the characterization of wheat glumes was implemented to perform automatically all the analysis procedures, reducing the execution time and contextually mistakes in the analysis process.

In order to reach the highest discrimination power, this macro was designed to compute 138 quantitative variables measured for each analysed left and right glume (Suppl. Info. 1 and 2). In particular, it was possible to measure 20 features descriptive of the glume surface colour and 18 parameters descriptive of the glume size and shape. Moreover, 78 quantitative Elliptic Fourier Descriptors (EFDs) were used to accurately describe the shape of the glume, as described by Orrù et al. (2013). Finally, the macro was kitted to compute 11 Haralick's descriptors including the relative standard deviations, as reported in Lo Bianco et al. (2015).

The 11 Haralick's descriptors measured on each glume to mathematically describe the surface texture and all the other morpho-colorimetric characters are available as supplementary information (Suppl. Info. 1 and 2).

2.3. Statistical analysis

Row data were submitted to one-way ANOVA and Tukey's was adopted as multiple comparison test. Percentage data were previously normalized with arcsine root square transformation.

The data, obtained from image analysis, were used to built a global database, including morpho-colorimetric, EFDs and Haralick's descriptors. Statistical elaborations were executed using SPSS software package release 16 (SPSS Inc. for Windows, Chicago, Illinois, USA), applying the same stepwise Linear Discriminant Analysis (LDA) algorithm suggested by Grillo et al. (2012). 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 (Holden et al., 2011).

The selection of the original features is carried out by a stepwise procedure. The stepwise method identifies and selects the most statistically significant features among them to use for the seed sample identification, 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 they are useful to describe what happens if a variable is inserted and removed, respectively, from the current model. This selective process starts with a model that does not include any of the original morpho-colorimetric features. At each step, the feature with the largest *F-to-enter* value that exceeds the entry criteria chosen ($F \geq 3.84$) is added to the model. The original features left out of the analysis at the last step have *F-to-enter* values smaller than 3.84, so no more are added. The process is automatically stopped when no remaining morpho-colorimetric features increased the discrimination ability (Grillo et al., 2012).

A cross-validation procedure was applied to verify the performance of the identification system, testing individual unknown cases and classifying them on the basis of all others (Gresta et al., 2016).

All the raw data were standardized before starting any statistical elaboration. Moreover, in order to evaluate the quality of the discriminant functions achieved for each statistical comparison, the Wilks' Lambda, the percentage of explained variance and the canonical correlation between the discriminant functions and the group membership, were computed. The Box's M test was executed to assess the homogeneity of covariance matrices of the features chosen by the stepwise LDA while the analysis of the standardized residuals was performed to verify the homoscedasticity of the variance of the dependent variables used to discriminate among the groups' membership (Box, 1949). Kolmogorov-Smirnov's test was performed to compare the empirical distribution of the discriminant functions with the relative cumulative distribution function of the reference probability distribution, while the Levene's test was executed to assess the equality of variances for the used discriminant functions calculated for groups membership (Levene, 1960).

To graphically highlight the differences among groups, multidimensional plots were drawn using the first three discriminant functions or, alternatively, when the number of discriminant groups n did not allow to obtain at least three discriminant functions ($n-1$), the two available discriminant functions and the

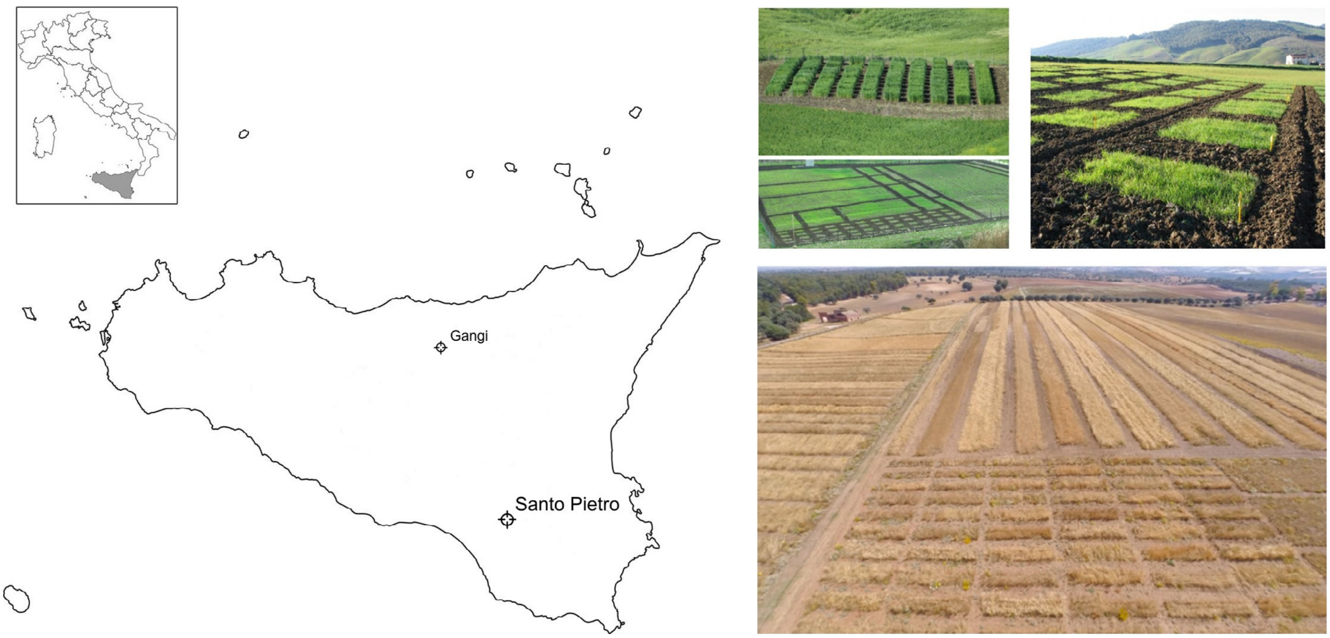

Fig. 1. Geografical location of Santo Pietro site and pictures of wheat field plots in different phenological phases.

Mahalanobis' square distance values were used (Mahalanobis, 1936).

3. Results

A preliminary statistical elaboration step was given on the basis of the current systematic classification. On this respect, all the nomenclatural classifications currently accepted, reported in the Wheat Genetic Resource Center of the Kansas State University web page (http://www.k-state.edu/wgrc/wheat-tax.html), were respected (Dorofeev et al., 1979; Gandilyan, 1980; Löve, 1984; Kimber and Feldman, 1987; Kimber and Sears, 1987; MacKey, 1988; van Slageren, 1994) for the three studied taxonomical entities, but for an easy reading the last published one was here considered (Goncharov, 2011), distinguishing among *T. aestivum* L., *T. durum* Desf. and *T. turgidum* L. The statistical comparison among the three botanical entities were able to reach a cross-validated correct identification of 100.0% (data not shown). The clear distinction among the groups is also highlighted by the 3D graphical representation of this comparison, drown using the Mahalanobis' square distance values together with the only two discriminant functions implemented by the stepwise LDA (Fig. 2A). Moreover, to graphically understand the normal distribution of the data used to compare the varietal groups, the homoscedasticity assessment of the variance of the used dependent variables were also conducted. Fig. 2B and C shows respectively, frequency and dispersion of the standardized residuals, while the Normal Probability Plot (P-P) reports the comparison between the cumulative probability expected and the observed one (Fig. 2D). The Kolmogorov-Smirnov normality test (*K-S*) was also executed to verify the normal distribution of the data, giving significance values lower than 0.05.

The first comparison, implemented among the five landraces of *T. aestivum* (Cuccitta [cuc], Maiorca di Pollina [mai1pol], Maiorca [mai6], Maiorcone [mce2] and Romano [rom2]), allowed to perfectly identify the glume samples, without giving misattributions among the tested landraces (Table 2).

Similarly, comparing the eight wheat landraces of *T. turgidum* (Bufala Bianca [bb2], Casedda [bivc], Bufala Nera Corta [bnc2], Bufala Nera Lunga [bnl1], Bufala Rossa Corta [brc2], Bufala Rossa Lunga [brl1], Ciciredda [cic1] and Paola [pao2]), a perfect cross-validated identification performance was reached, in spite of the reduced glume sample amount (Table 3).

In order to assess the discrimination power of the implemented statistical system, also for the 39 landraces of durum wheat, a third comparative analysis was conducted. In this case, an overall percentage of correct identification of 89.7% was achieved (data not shown), with performances ranged between 71.1% (Margherito 02 [mar2]) and 100.0% (Capeiti [cat], Cappelli [cal], Castiglione Glabro [cas3gl], Martinella [mla1], Semenzella [sem1] and Trentino [tre2]). Main misattributions were recorded for the landraces Bidì [bd3] and Margherito 02 [mar2], erroneously classifying for Margherito 06 [mar6] the 12.9% and 17.8% of the cases, respectively (data not shown). Moreover, the landrace Gioia [gio1] was mainly misidentified for Castiglione Glabro [cas3gl] in 11.6% of the cases (data not shown). Other little misidentifications were recorded between the landraces Timilia [tim] and Tumminia SG3 [tumsg3], and between the landraces Russello [rus1] and Russello 13 SG8 [russg8]. Not particularly significant mistakes were revealed for the landrace Chiattulidda [chi1], exclusively misattributed for Scavuzza [sca1] in 16.7% of the cases, and for the variety Biancuccia [bia1], correctly identified in 75.5% of the cases but mainly misattributed to Sicilia [sic1] in 13.3% of the cases (data not shown).

Finally, a comparative analysis, including all the studied landraces together, was done to assess the system capability to discriminate the wheat landraces regardless of the systematic classification. Fig. 3 shows the 3D graphical representation of the group centroids, only distinguishing in colour the partnership to different systematic groups, for an easier reading. In this case the overall performance of correct identification reaches the 93.7%, with little misattributions reflecting the percentages reported for the comparisons cited above (data not shown).

In Fig. 4, the glume samples of some of the studied landraces are reported.

After having assessed the actual identification power of the statistical system based on glume morpho-colorimetric features, a validation test was conducted adding into the system an unknown

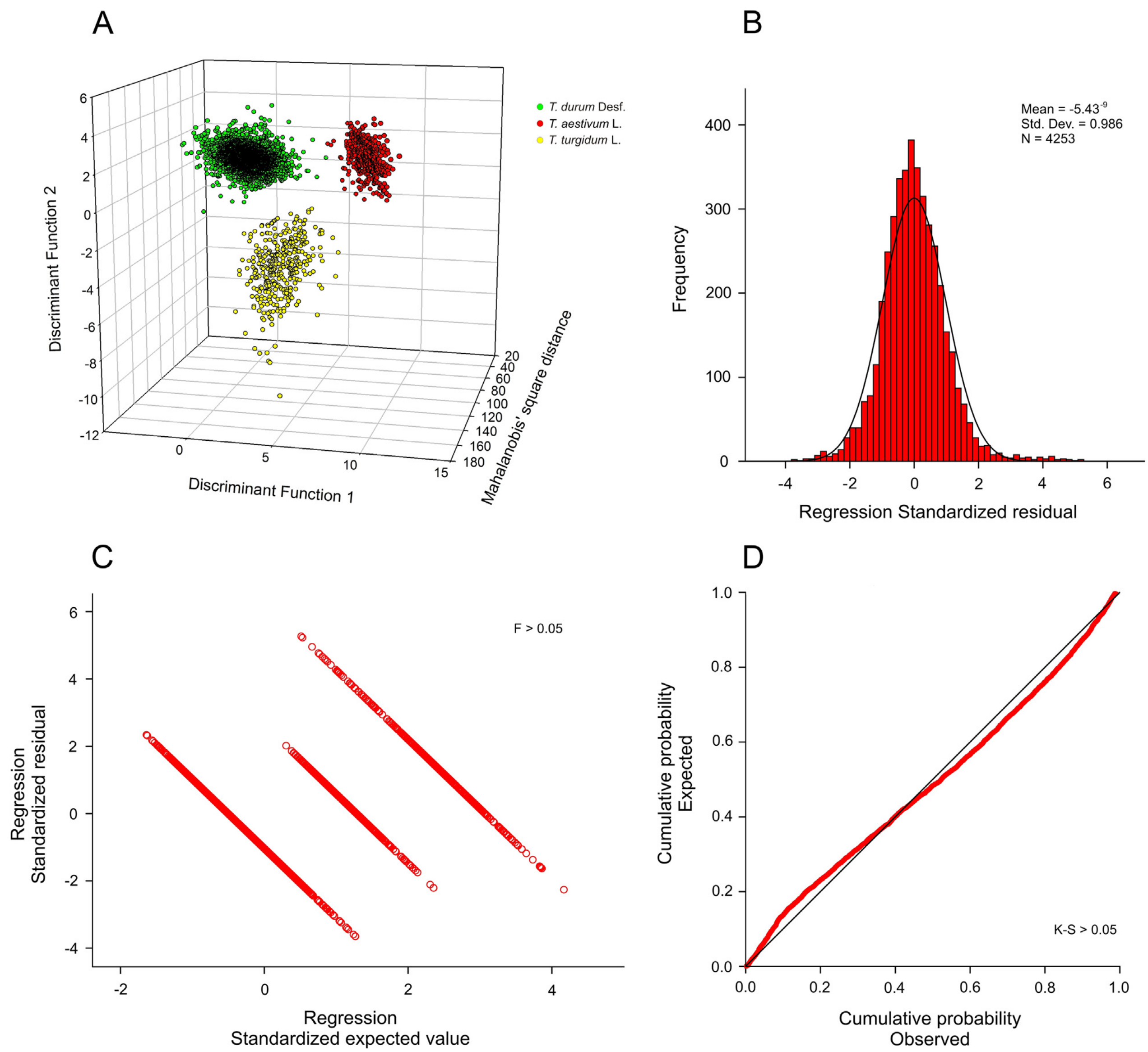

Fig. 2. (A) Graphical representation of the discriminant scores of the three studied botanical entities of the genus *Triticum*; (B) histogram of the standardised residuals; (C) dispersion plot of the standardised residuals tested with Levene's test (*F*); (D) normal probability plot (*P-P*) tested with Kolmogorov-Smirnov's test (*K-S*).

Table 2
Percentage of correct identification among varieties belonging to the *T. aestivum* L. species. In parentheses, number of seeds analysed. Bold values indicate the correct identification performance.

	cuc1	mai1pol	mai6	mce2	rom2	Tot
cuc1	**100.0% (99)**	–	–	–	–	100.0% (99)
mai1pol	–	**100.0% (84)**	–	–	–	100.0% (84)
mai6	–	–	**100.0% (76)**	–	–	100.0% (76)
mce2	–	–	–	**100.0% (76)**	–	100.0% (76)
rom2	–	–	–	–	**100.0% (89)**	100.0% (89)
Overall						**100.0% (424)**

glume sample [UGS], in order to allow its identification and test and validate the system. The 54 unknown glumes were entirely identified as Vallelunga (data not shown).

In the evaluation of the parameters that more than other influenced the discrimination process of the studied wheat varieties, the most important variables chosen by the stepwise LDA were related both to glume shape and surface colour. In Table 4 the best five variables used by the system are reported. Although the LDA was able to reach a very high percentage of correct identification, the whole discriminant analysis had needed of 83 over the 138 measured variables to discriminate among the varieties, completing the discrimination process in 95 consecutive steps. Globally, 21 densitometric features descriptive of the seed surface colour and textural, 13 morphological parameters descriptive of seed size and contour shape, and 49 Elliptic Fourier Descriptors, were statistically selected and used by the LDA (data non shown).

Table 3
Percentage of correct identification among varieties belonging to the *T. turgidum* L. In parentheses, number of seeds analysed. Bold values indicate the correct identification performance.

	bb2	bivc	bnc2	bnl1	brc-b1	brl1	cic1	pao2	Tot
bb2	**100.0% (45)**	–	–	–	–	–	–	–	100.0% (45)
bivc	–	**100.0% (24)**	–	–	–	–	–	–	100.0% (24)
bnc2	–	–	**100.0% (28)**	–	–	–	–	–	100.0% (28)
bnl1	–	–	–	**100.0% (35)**	–	–	–	–	100.0% (35)
brc-b1	–	–	–	–	**100.0% (40)**	–	–	–	100.0% (40)
brl1	–	–	–	–	–	**100.0% (92)**	–	–	100.0% (92)
cic1	–	–	–	–	–	–	**100.0% (31)**	–	100.0% (31)
pao2	–	–	–	–	–	–	–	**100.0% (93)**	100.0% (93)
Overall									**100.0% (388)**

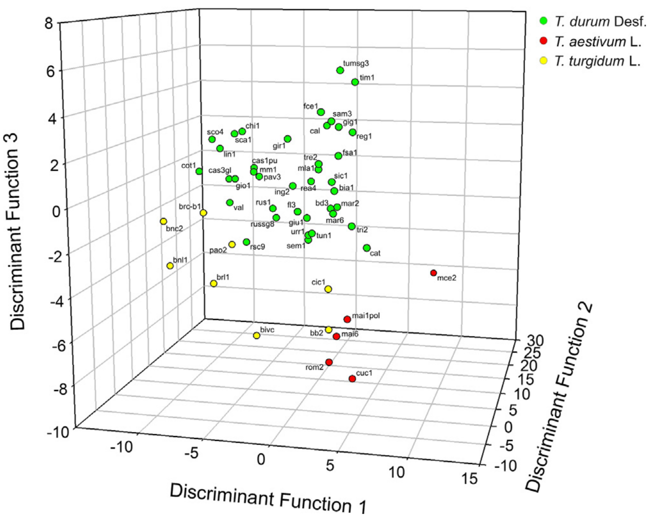

Fig. 3. Graphical representation of the discriminant scores of the group centroids, for all the investigated wheat landraces. Different colours indicate the partnership to different systematic groups (green: *T. durum* Desf.; red: *T. aestivum* L.; yellow: *T. turgidum* L.). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

Although the studied landraces, belonging to the species *T. durum* and *T. turgidum*, may be grouped because both naked tetraploids belonging to the same *Dicoccoides* Flaksb. section, a differentiation between them was adopted due to their marked morphological differences. Even though the three botanical entities resulted perfectly distinguishable on the basis of the glume morphology, this preliminary comparison was useful to facilitate the discrimination among the wheat landraces.

The comparisons among the five landraces of *T. aestivum* and among the eight landraces of *T. turgidum* proved the absolute effectiveness of the system, although a highest number of glume samples for each landrace should increase the statistical significance of the results.

Good identification performance was achieved also from the comparative analysis among the durum wheat landraces, although some little but significant misattributions testify the efficacy of the method. It revealed some plausible similarities among the landraces [bd3], [mar2] and [mar6]. Instead, the two landraces Bidì and Margherito have origin from the N-African population "Jean Rètifah". In particular, Bidì (line 74) and Margherito were independently selected, by genealogical selection (pure line), by Tucci (University of Palermo) (De Cillis, 1935) and Santagati (University of Catania), respectively. The name "Margherito" derives from the town district of Ramacca (Catania) were, for the first time, this landrace was tested (Prestianni, 1926). Nevertheless, probably this population was already previously classified as "AP[4]" by the Tunisian Botanical Service and known in Tunisia as "Mahmoudi" (De Cillis, 1939). These landraces, considered as unique by Venora and Blangiforti (2017) show differences form the phenological point of view, specially for the precocity, characterizing and spreading them in different areas of the island. Bidì, being later in growth, spread in hill and high hill; while Margherito, as earlier in growth, easily spread in plain and low hill. Moreover, the samples [mar2] and [mar6] used for this study, are two different accessions of the same landrace Margherito. An important consideration deserves "Cappelli", another genealogical selection from the same N-African population, registered by Strampelli in 1915 and only recently spread in Sicily. Even though Cappelli shows very similar morphological and biological characters to Bidì and Margherito, and in spite of the reduced amount of analysed glumes, it was perfectly identified by the system. It is probably due to the narrow genetic variability of the original selection of Strampelli, at that time done following the severe protocol required to register the variety.

Also the misattribution of the landrace [gio1] for [cas3gl] have genealogical explanation. Indeed, it seems that the landrace Gioia, described for the first time by De Cillis (1942), is a selection of the landrace Castiglione, one of the most ancient Sicilian durum wheat varieties, whose first historical note is dated back to the beginning of the XIX century (Venora and Blangiforti, 2017). Moreover, the both these landraces are historically spread and cultivated into the same areas of the Sicilian backcountry, between the provinces of Palermo, Agrigento and Enna.

Similar consideration is relevant for the misclassifications revealed among [tim] and [tumSG3], and among [rus1] and [russg8]. Tumminia SG3 is indeed the last intra-population selection of the landrace Timilia, after "Timilia SG1" with black awns and "Timilia SG2" with white awns, derived by the genetic improvement programs conducted by De Cillis during the 30′s (Venora and Blangiforti, 2017). Likewise, Russello SG8 is the last intra-population of the landrace Russello, after "Russello SG7" selected by De Cillis during his experimentations. Tumminia SG3 and Russello SG8 were both recorded at the Community Plant Variety Office by the Stazione Sperimentale di Granicoltura per la Sicilia, in 2007.

Regarding the misattribution percentages recorded for the two landraces Chiattulidda and Biancuccia, partially identified as Scavuzza and Sicilia, respectively, it is important to highlight the reduced amount of glume samples in both the cases. For this study, only 30 glumes of the landrace [chi1] and 40 glumes of the sample [bia1] were available. This, together with the reduced peculiar phenotipical characters of the glumes of these landraces, is the reason of these misidentifications that should have be considered not significant.

From the last comparison, implemented among all the studied landraces without distinguishing the systematic partnership, the system preserved its identification capability, increasing the over-

Fig. 4. Representative glume samples of some of the landraces considered in the study.

Table 4
The best five variables over the 83 selected by the LDA for glumes identification. The number of steps, feature name (according to the Supplemental Table S2), *F-to-remove* and the *Tolerance* values are reported.

Step	Feature	*F-to-Remove*	*Tolerance*
1	FD_9	182.395	0.270
2	FD_{13}	130.696	0.201
3	FD_{18}	113.424	0.328
4	L_{Mean}	110.144	0.023
5	R_{sd}	86.904	0.012

all percentage respect to the comparison exclusively conducted among the durum wheat landraces. This was due to the high performances achieved for the two other systematic groups (*T. turgidum* and *T. aestivum*). In Fig. 3, the systematic groups are highlighted and inside each, some relevant little groupings, genetically or genealogically related, are identifiable. For instance, the four turgidum Bufala [brc-b1], [bnc2], [bnl1] and [brl1] were very closely collocated; as well as the two Maiorca [mai1pol] and [mai6] and Maiorcone [mce2], and many durum wheat landraces: [mar2], [mar6] and [bd3]; [gio1] and [cas3gl]; [rus1] and [russg8]; or [tim1] and [tumsg3].

The last analysis was conducted in order to try the identification of the unknown glume sample [UGS] from Madonie mountains, named "Nivuru". This analysis allowed to test and validate the effectiveness of the identification system, univocally identifying all the unknown glumes as Vallelunga, without doubts or little uncertainties, in spite of the high within variability of wheat landraces or populations. This result is in accordance with Venora and Blangiforti (2017), explaining that the landrace Vallelunga is commonly named, in some regional areas "Regina Sammartinara" (De Cillis, 1942), while in others "Nivuru", although no official documentation exists about its origin.

Finally, considering the very good identification performance recorded for each conducted comparison, in function of the high variability included in the samples, derived from the four consecutive years of reaping, it is appropriate to highlight the stability of the glume morpho-colorimetric characters for identification scopes.

5. Conclusion

This work represent the first attempt of wheat landraces identification based on glume phenotypic characters, applying image analysis techniques. The achieved results here discussed allowed to demonstrate the usefulness of this discrimination system for the identification and classification wheat landraces, notoriously very difficult to do. The technique here proposed, conveniently sustained by a conspicuous database, can be undoubtedly considered a helpful identification tool both for commercial varieties and for no genetically defined samples, such as populations or landraces.

Considering the growing interest in local old wheat landraces, strongly linked to the renewed appreciation in traditional and typical local products, the obtained results support the application of the image analysis system not only for grading purposes, but also to define the product traceability, in order to get a "market card" for wheat landraces. Food traceability is becoming increasingly relevant, especially in terms of international trade. For the export and import of food, the development of traceability systems has been identified as a priority, mainly in connection with food safety.

Considering the heterogeneous nature of the wheat landrace samples used in this study, in order to validate these preliminary achievements, further trials will have to be conducted focusing on the collection of new data, enriching the database with new and accurate information, allowing to the system to give results more and more reliable.

Acknowledgements

The authors thank to Rosario Failla for the great contribution in field work and management of wheat samples in the germplasm bank of the Stazione Sperimentale di Granicoltura per la Sicilia, and to Concetta Ravalli for the support in image acquisition.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.compag.2017.07.024.

References

Arefi, A., Modarres Motlagh, A., Farrokhi Teimourlou, R., 2011. Wheat class identification using computer vision system and artificial neural networks. Int. Agrophys. 25, 319–325.
Barker, D.A., Vouri, T.A., Hegedus, M.R., Myers, D.G., 1992. The use of Fourier descriptors for the discrimination of Australian wheat varieties. Plant Variet. Seeds 5, 93–102.
Box, G.E.P., 1949. A general distribution theory for a class of likelihood criteria. Biometrika 36, 317–346.
Camacho, V., Taina, C., Maxted, N., Scholten, M., Ford-Lloyd, B., 2005. Defining and identifying crop landraces. Plant Gen. Resour. 3, 373–384.
Cebola Lidon, F., Almeida, A.S., Leitão, A.L., Silva, M.M., Pinheiro, N., Maçãs, B., Costa, R., 2014. A synoptic overview of durum wheat production in the Mediterranean region and processing following the European Union requirements. Emirates J. Food Agric. 26 (8), 693–705.
Chaugule, A.A., Mali, S.N., 2016. Identification of paddy varieties based on novel seed angle features. Comput. Electron. Agric. 123, 415–422.
De Cillis, E., 1935. La concimazione dei frumenti a maturazione precoce e normale, guida pratica per l'Italia meridionale e insulare. Comitato Nazionale Incremento Concimazioni, Roma.
De Cillis, U., 1939. Fattori di progresso della granicoltura siciliana: varietà elette di grano e loro produzione. Palermo, (Enciclopedia Agraria Italiana - UTET - 1985pp. 605), pp. 8–12.
De Cillis, U., 1942. I frumenti siciliani. Stazione Sperimentale di Granicoltura per la Sicilia, Catania. Pubbl. 9, 1942.
Dorofeev, V.F., Filatenko, A.A., Migushova, E.F., Udaczin, R.A., Jakubziner, M.M., 1979. Wheat. In: Dorofeev, V.F., Korovina, O.N. (Eds.), Flora of Cultivated Plants, vol. 1. Kolos, Leningrad (St. Petersburg), Russia (in Russian), p. 346.
FAO web page (http://www.fao.org/faostat/en/#home) (accessed 02.08.17).
Frankel, O.H., 1970. Evaluation and utilisation. Introduction remarks. In: Frankel, O. H., Bennett, E. (Eds.), Genetic Resources in Plants. Their Exploration and Conservation. Blackwell, Oxford and Edinburgh, pp. 395–402.
Fujita, Y., Fukuoka, H., Yano, H., 2009. Identification of wheat cultivars using ESTSSR markers. Breed. Sci. 59 (2), 159–167.
Gallo, G., Porfiri, O., Venora, G., Lombardo, G.M., Blangiforti, S., 2004. I frumenti siciliani: patrimonio da mantenere e valorizzare. Maimone, Catania.
Gandilyan, P.A., 1980. Key to Wheat, Aegilops, Rye and Barley. Academy of Science, Armenian SSR, Erevan, p. 190.
Giancaspro, A., Colasuonno, P., Zito, D., Blanco, A., Pasqualone, A., Gadalet, A., 2016. Varietal traceability of bread 'Pane Nero di Castelvetrano' by denaturing high pressure liquid chromatography analysis of single nucleotide polymorphisms. Food Control 59, 809–817.
Goncharov, N.P., 2011. Genus *Triticum* L. taxonomy: the present and the future. Plant Syst. Evol. 295, 1–11.
Gresta, F., Santonoceto, C., Ceravolo, G., Formantici, C., Grillo, O., Ravalli, C., Venora, G., 2016. Productive, qualitative and seed image analysis traits of guar (*Cyamopsis tetragonoloba* (L.) Taub). Aust. J. Corp Sci. 10 (7), 1052–1060.
Grillo, O., Blangiforti, S., Venora, G., 2016. Traditional food with authentic raw material: the Sicilian wheat landraces. In: Proceedings of the "Bonding tradition with innovation: successful strategies in food chain value", 12/13 September 2016, Olsztyn (Poland).
Grillo, O., Draper, D., Venora, G., Martínez-Laborde, J.B., 2012. Seed image analysis and taxonomy of *Diplotaxis* DC. (Brassicaceae Brassiceae). Syst. Biodivers. 10 (1), 57–70.
Guarda, G., Padovan, S., Delogu, G., 2004. Grain yield, nitrogen-use efficiency and baking quality of old and modern Italian bread-wheat cultivars grown at different nitrogen levels. Eur. J. Agron. 21, 181–192.
Guevara-Hernandez, F., Gomez-Gil, J., 2011. A machine vision system for classification of wheat and barley grain kernels. Spanish J. Agric. Res. 9 (3), 672–680.
Haralick, R.M., Shanmugam, K., Its'hak, D., 1973. Textural features for image classification. IEEE Trans. Syst. Man Cybernet. 3, 610–621.
Hâruta, O., 2011. Elliptic Fourier analysis of crown shapes in *Quercus petraea* trees. Ann. Forest Res. 54, 99–117.
Holden, J.E., Finch, W.H., Kelley, K., 2011. A comparison of two-group classification methods. Educ. Psychol. Meas. 71, 870–901.
Jones, H., Lister, D.L., Bower, M.A., Leigh, F.J., Smith, L.M., Jones, M.K., 2008. Approaches and constraints of using existing landrace material to understand agricultural spread in prehistory. Plant Gen. Resour. 6, 98–112.
Keefe, P.D., Draper, S.R., 1986. The measurement of new characters for cultivar identification in wheat using machine vision. Seed Sci. Technol. 14, 715–724.
Kimber, G., Feldman, M., 1987. Wild Wheat: An Introduction Special Report 353. College of Agriculture, University of Missouri, Columbia, p. 146.
Kimber, G., Sears, E.R., 1987. Evolution in the genus Triticum and the origin of cultivated wheat. In: Heyne, E.G. (Ed.), Wheat and Wheat Improvement. second ed. American Society of Agronomy, Madison, WI, pp. 154–164.
Landjeva, S., Ganeva, G., Korzun, V., Palejev, D., Chebotar, S., Kudrjavtsev, A., 2015. Genetic diversity of old bread wheat germplasm from the Black Sea region evaluated by microsatellites and agronomic traits. Plant Gen. Resour. 13 (2), 119–130.
Levene, H., 1960. Robust tests for equality of variances. In: Olkin, I., Ghurye, S.G., Hoeffding, W., Madow, W.G., Mann, H.B. (Eds.), Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press, Palo Alto, CA, USA, pp. 278–292.
Lo Bianco, M., Grillo, O., Sarigu, M., Cremonini, R., Venora, G., 2015. Characterisation of Italian bean landraces (*Phaseolus vulgaris* L.) using seed image analysis and texture descriptors. Aust. J. Crop Sci. 9 (11), 1022–1034.
Lo Bianco, M., Siracusa, L., Dattilo, S., Venora, G., Ruberto, G., 2017. Phenolic fingerprint of Sicilian modern varieties and old durum wheat landraces: a tool to assess with diversity. Cereal Chem. (submitted for publicattion).
Lombardo, G.M., 2004. Evoluzione varietale del frumento duro in Sicilia nel corso del XX secolo. In: Maimone, G. (Ed.), I frumenti Siciliani – Un patrimonio da mantenere e valorizzare. A cura della Stazione Sperimentale di Granicoltura per la Sicilia, Catania.
Löve, Á., 1984. Conspectus of the Triticeae. Feddes Repert 95 (7–8), 425–521.
MacKey, J., 1988. A plant breeder's perspective on taxonomy of cultivated plants. Biol. Zentralblatt 107, 369–379.
Mahalanobis, P.C., 1936. On the generalised distance in statistics. Proc. Natl. Instit. Sci. India 12, 49–55.
Mebatsion, H.K., Paliwal, J., Jayas, D.S., 2013. Automatic classification of non-touching cereal grains in digital images using limited morphological and color features. Comput. Electron. Agric. 90, 99–105.
Migliorini, P., Spagnolo, S., Torri, L., Arnoulet, M., Lazzerini, G., Ceccarelli, S., 2016. Agronomic and quality characteristics of old, modern and mixture wheat varieties and landraces for organic bread chain in diverse environments of northern Italy. Eur. J. Agron. 79, 131–141.
Morancho, J., 2000. Production et commercialization du blé dur dans le monde, In: Di Fonzo, N., Kaan, F., Nachit, M., (Eds.). Durum Wheat Improvement in Mediterranean Region", New Challenges, CIHEAM, Zaragoza. Options Méditerranéennes A. 40, pp. 19–34.
Newton, A.C., Akar, T., Baresel, J.P., Bebeli, P.J., Bettencourt, E., Bladenopoulos, K.V., Czembor, J.H., Fasoula, D.A., Katsiotis, A., Koutis, K., Koutsika-Sotiriou, M., Kovacs, G., Larsson, H., De Carvalho, M.A.A.P., Rubiales, D., Russell, J., DosSantos, T.M.M., Vaz Patto, M.C., 2010. Cereal landraces for sustainable agriculture. Agron. Sustain. Dev. 30, 237–269.

Orrù, M., Grillo, O., Lovicu, G., Venora, G., Bacchetta, G., 2013. Morphological characterization of *Vitis vinifera* L. seeds by image analysis and comparison with archaeological remains. Vegetation History Archaeobotany 22, 231–242.
Pasqualone, A., Delvecchio, L.N., Mangini, G., Taranto, F., Blanco, A., 2014. Variability of total soluble phenolic compounds and antioxidant activity in a collection of tetraploid wheat. Agric. Food Sci. 23 (4), 307–316.
Pasqualone, A., Lotti, C., Blanco, A., 1999. Identification of durum wheat cultivars and monovarietal semolinas by analysis of DNA microsatellites. Eur. Food Res. Technol. 210, 144–147.
Pasqualone, A., Lotti, C., Bruno, A., De Vita, P., Di Fonzo, N., Blanco, A., 2000. Use of ISSR markers for cultivar identification in durum wheat. Opt. Mediterraneennes 40, 157–161.
Peruffo, A.D.B., Pogna, N.E., Pallavicini, C., Pegoraro, E., 1985. Diagrammi elettroforetici delle gliadine e chiave di identificazione delle varietà di grano tenero iscritte nel registro delle varietà. Agricoltura e Ricerca 11, 75–81.
Pourreza, A., Pourreza, H., Abbaspour-Fard, M.H., Sadrnia, H., 2011. Identification of nine Iranian wheat seed varieties by textural analysis with image processing. Comput. Electron. Agric. 83, 102–108.
Prestianni, N., 1926. Frumenti Coltivati in Sicilia. Commissione Provinciale di Caltanissetta, Caltanissetta.
Sciacca, F., Cambrea, M., Licciardello, S., Pesce, A., Romano, E., Spina, A., Virzì, N., Palumbo, M., 2014. In: Porceddu, E., Damania, A.B., Qualset, C.O. (Eds.), Evolution of Durum Wheat From Sicilian Landraces to Improved Varieties. Proceedings of the International Symposium on Genetics and Breeding of Durum Wheat. CIHEAM, Bari, pp. 139–145 (Options Méditerranéennes : Série A. Séminaires Méditerranéens; n. 110). International Symposium : Genetics and Breeding of Durum Wheat, 2013/05/27-30, Rome (Italy).
Shahin, M.A., Symons, S.J., 2003. Colour calibration of scanners for scanner independent grain grading. Cereal Chem. 80, 285–289.
van Slageren, M.W., 1994. Wild wheats: a monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae). Wageningen Agriculture University Papers 1994(7), pp. 513.
Smykalova, I., Grillo, O., Bjelkova, M., Hybl, M., Venora, G., 2011. Morpho-colorimetric traits of *Pisum* seeds measured by an image analysis system. Seed Sci. Technol. 39, 612–626.
Smykalova, I., Grillo, O., Bjelkova, M., Pavelek, M., Venora, G., 2013. Phenotypic evaluation of flax seeds by image analysis. Ind. Crops Prod. 47, 232–238.
Spina, A., Blangiforti, S., Cambrea, M., Gallo, G., Licciardello, S., 2008. Valorizzazione di germoplasma locale di frumento duro per la produzione di pani tipici. In: Proceedings of the VIII Convegno Nazionale ''La biodiversità – una risorsa per sistemi multifunzionali", 21/23 April 2008, Lecce (Italy).
SPSS, 2007. Application Guide, SPSS Version 16.0. SPSS Inc, Chicago.
Sugiyama, M., 2007. Dimensionality reduction of multimodal labeled data by local Fisher Discriminant Analysis. J. Mach. Learn. Res. 8, 1027–1061.
Szczypiński, P.M., Klepaczko, A., Zapotoczny, P., 2015. Identifying barley varieties by computer vision. Comput. Electron. Agric. 110, 1–8.
Szczypiński, P.M., Zapotoczny, P., 2012. Computer vision algorithm for barley kernel identification, orientation estimation and surface structure assessment. Comput. Electron. Agric. 87, 32–38.
Vavilov, N.I., 1957. World Resources of Cereals, Leguminous Seed Crops and Flax and Their Utilization in Plant Breeding. Academy of science of the URSS, Moscow.
Venora, G., Blangiforti, S., 2017. I grani antichi siciliani. Manuale tecnico per il riconoscimento delle varietà locali dei frumenti siciliani. Le Fate Editore, Ragusa.
Venora, G., Grillo, O., Saccone, R., 2009. Quality assessment of durum wheat storage centers in Sicily: evaluation of vitreous, starchy and shrunken kernels by image analysis system. J. Cereal Sci. 49, 429–440.
Zapotoczny, P., 2011. Discrimination of wheat grain varieties using image analysis: morphological features. Eur. Food Res. Technol. 233, 769–779.
Zayas, I., Pomeranz, Y., Lai, F.S., 1989. Discrimination of wheat and non-wheat components in grain samples by image analysis. Cereal Chem. 66, 233–237.
Zeven, A.C., 1998. Landraces: A review of definitions and classifications. Euphytica 104 (127–139), 1998.
Wheat Genetic Resource Center of the Kansas State University web page (accessed 06.04.17).