Australian Journal of Crop Science
AJCS
AJCS 9(11):1022-1034 (2015) ISSN:1835-2707

Characterisation of Italian bean landraces (*Phaseolus vulgaris* L.) using seed image analysis and texture descriptors

Marisol Lo Bianco[1], Oscar Grillo[1]*, Roberto Cremonini[2], Marco Sarigu[3], Gianfranco Venora[1]

[1]Stazione Consorziale Sperimentale di Granicoltura per la Sicilia. Via Sirio, 1, 95041 Borgo Santo Pietro - Caltagirone (CT), Italy
[2]Dipartimento di Biologia, Università di Pisa. Via Ghini, 5, I-56126, Pisa, Italy
[3]Centro Conservazione Biodiversità (CCB), Dipartimento di Scienze della Vita e dell'Ambiente (DISVA), Università degli Studi di Cagliari, V. le S. Ignazio da Laconi 11-13, 09123 Cagliari, Italy

***Corresponding author:oscar.grillo.mail@gmail.com**

Abstract

Seed morphological traits were used to identify 67 Italian bean (*Phaseolus vulgaris* L.) accessions, belonging to 58 Italian landraces. An overall of 138 size, shape and texture descriptors were measured, on each seed, using image analysis techniques. The achieved data, analysed applying the stepwise Linear Discriminant Analysis, allowed to discriminate among bean landraces, also identifying the harvest year and the cropping areas. Comparative analyses were carried out to verify the possibility to distinguish seeds belonging to the same landrace but grown applying different agricultural practices. Preliminarily, it was possible to discriminate three main color categories of bean seeds, with an overall performance of 99.1%. Moreover, for each of these three categories, the belonging bean landraces were identified, with overall correct identification percentages included between 94.3% and 99.7%. Following the same procedure, it was possible to assess the possibility to identify the bean landraces origin, reaching overall correct identification percentage higher than 88%. Also considering the effect of the cropping year, the cultivation region and the agricultural practices, high identification performances were recorded. The results support the application of the computer vision system not only for the identification, classification or grading purpose, but also to define the product traceability, in order to get a "market card" for landrace beans.

Keywords: Computer vision; EFDs; Haralick's features; LDA; Seed morphology; Traceability.
Abbreviations: EFDs-Elliptic Fourier Descriptors.

Introduction

Common bean (*Phaseolus vulgaris* L.) is one of the most important grain legume for direct human consumption, since it represents a cheap source of dietary proteins. After its introduction from the Americas in the 16th century, it promptly reached a wide diffusion in Europe (Piergiovanni and Lioi, 2010), where farmers selected, for different morpho-productive traits, a large amount of landraces (Rodiño et al., 2009). The selective pressure, due to an adaptive evolutionary process, in addition to the microclimate of cultivation areas, as well as local constraints to production and different consumer preferences, resulted in a wide differentiation of landraces that can be observed within the European common bean germplasm. These landraces represent local specialties very appreciated for their taste, high nutritional value, short cooking time, thin coat and good yield (Piergiovanni et al., 2000). Nevertheless, some of them have a great economical potential, especially as quality food produced under low input agro-systems (Negri, 2003). They generally have local names, identifying the well established geographical area. Some of them, have recently obtained the European trademark as PGI (Protected Geographical Indication), one of the quality recognition standards introduced by the European Community (CEE regulations n. 2081/92 passed by the European Council on 14/7/92). The punctual distribution of these local varieties and the following great assortment of assigned names, contributed to enrich the bean varietal heritage (Piergiovanni and Laghetti, 1999). Consequently, the current whole amount of landraces representative of the Italian territory is just less than 150, sometimes reported as accessions (Logozzo et al., 2006; Reggi et al., 2013). Several techniques, involving the analysis of morphological, biochemical and molecular markers (Lioi et al., 2005; Sicard et al., 2005; Grisi et al., 2007; Marotti et al., 2007; Mercati et al., 2012; Reggi et al., 2013), can be used to identify the germplasm collections and assess the genetic relationships among accessions within a species and among biotypes of a same landraces (De La Fuente et al., 2012; Diniz et al., 2014)). These methodologies may be used to trace and authenticate food and products, improving safety and quality. Nevertheless, technologies and costs of genotyping and phenotyping can be currently too expensive, labor intensive and environmentally sensitive. Over the next two decades, the development of phenotyping strategies will almost certainly mirror innovations in genotyping technology that have occurred over the last 20 years, characterized by increasing automation and throughput. As the science of phenotyping evolves, emphasis will increasingly be placed on generating information that is as accurate (able to effectively measure traits and/or performance characteristics), precise (small variance associated with replicated measurement), and as relevant as possible, while keeping costs within reasonable limits (Houle et al., 2010; Cobb et al., 2013). Therefore, at the current status, considering the molecular studies on finding specific markers to distinguish particular landraces and cultivars, it is possible to assert that phenotyping by computer vision is a least expensive method and equally efficient in term of distinctiveness, with high potentialities to be considered, in the near future, as the main technique for the characterization and taxonomic identification (Dreher et al., 2003; Park et al., 2009; Orru et al., 2012). Many studies have been conducted to distinguish different agricultural products on the basis of shape, size and color, using image analysis systems (Mahajan et al., 2015). This technique was applied to the morphological and textural characterization of many commercial types of grains and seeds, such as lentils (Shahin and Symons, 2003; Venora et al., 2007), peas (Smykalova et al., 2011), vetch (Grillo et al., 2011), flax (Smykalova et al., 2013), grapevine (Orrù et al., 2012), in order to discriminate among varieties and/or commercial categories (Venora et al., 2009a), as well as to characterize and discriminate among wild seeds belonging to various taxonomical ranks (Bacchetta et al., 2008, 2010; Grillo et al., 2010, 2012, 2013). Currently, chromatic and geometrical measurements can be successfully carried out in a precise, accurate and repeatable way, giving objective information (Granitto et al., 2003; Venora et al., 2007; Bacchetta et al., 2008). In a previous work, Venora et al. (2009b) developed a macro program, based on image analysis, able to identify 15 Italian common bean landraces on the basis of 26 quantitative morpho-colorimetric variables of whole seed surface and their spots. The possibility to differentiate beans by their harvest year and/or cultivation regions was demonstrated for the first time, giving in some extent, a product traceability.

According to the achievements published on recent papers (Grillo et al., 2010; Pinna et al., 2014; Lo Bianco et al.,), the discrimination power of an identification system not only depends on the intra-specific representativeness of the analyzed samples, but also on the quality and quantity of the parameters measured and used to discriminate among groups, as well as on the dimension and variability degree of the groups. For this reason, it is plausible that an increase in measured features and in seed amount for each landrace-class, could be useful to improve the identification performance reached by Venora et al. (2009b).

The recent literature proves that features descriptive of seed surface texture, as well as of its geometric shape, seems to be strongly discriminant parameters (Diamond et al., 2004; Gerger and Smolle, 2004; Iwata et al., 2002, 2004; Kawabata et al., 2009; Nanni et al., 2010). Computing the Haralick's texture indicators, able to quantitatively measure the color tones variation within a surface, so defining the real chromatic pattern; and the Elliptic Fourier Descriptors (EFDs), able to accurately define the shape of the bi-dimensional profile of a seed projection; seed texture and shape can be carefully defined. The aims of present work are: (1) to implement a statistical classifier, based on seed morpho-colorimetric features, including Elliptic Fourier (EFDs) and Haralick's descriptors, able to identify bean landraces; (2) to validate the statistical identification system with the data of 67 Italian bean accessions representative of the Italian territory; (3) to assess the possible differences in the same landrace grown in different cultivation regions or tilled in different agricultural systems.

Results and discussion

Comparisons among differet coat color beans

A preliminary statistical elaboration step was given on the basis of seed coat main color: white, mono-colored (including landraces with a single seeds coat color, from cream to dark brown) and bi-colored or spotted seed coat (Fig. 1). Applying this discrimination model, percentages of correct identification, ranged between 98.2% (mono-colored beans) and 100.0% (white beans), were reached, with an overall performance of 99.1%, confirming the importance of the color features for the bean discrimination (data not shown) . Although these three categories are easily distinguishable also by visual inspection, this comparison was necessary both to validate the system, and to fix major categories to deeply investigate with further comparisons. Afterwards, a classifier was developed for each of the three main color categories of bean seed coat, only considering the landrace name as grouping variable. In table 1, the classifiers cross-validated performances are given, for the white and mono-colored bean landrace groups, respectively. Relating to the 13 white landraces, the overall percentage of correct classification was 96.0%, the lowest was recorded for Cannellino di Pisa [CaP] (85.8%) and the highest for Triverde [Tri] (100.0%). Whereas, the eight mono-colored coat bean landraces reached an overall correct classification of 99.7%, recording for Giallo [Gia], Moitano [Moi], Tabacchino [Tab] and Vellutina di Ragusa [Vel] the perfect identification rate of 100.0%. In both statistical elaborations, the mean seed weight represented the more powerful parameter of the discriminant functions, showing a significantly high value of *F-to-remove* (data not shown). This feature was followed, for the white bean classification, by several shape descriptive parameters, explaining the wide variability of seed sizes. Regard to mono-colored coat beans, after the mean seed weight, the most important parameters, chosen by the stepwise LDA, were related to color and textural information, with a particular focus on the Entropy, that represents the variability degree of the surface color, proving the power of this kind of features in the discrimination process (data not shown).

Table 2 shows the percentages of correct identification reached for the bi-colored coat bean accessions. The overall performance of 94.3% was achieved. In this group, 23 out of 37 accessions were distinguished above the 90%; for nine accessions, a correct classification range included between 82.9% (Panzaredda Nera [PaN], misclassified as Badda Niura [BaN] in the 6.5% of cases, and as Mussuniuru [Mus] in the 5.9%) and 89.5% (Giovanna [Gio], misclassified as San Michele [SaM] in the 7.9% of cases) was recorded. Only five bi-colored coat bean landraces (Billò [Bil], Borlotto Bianco [BoB], Fiumara [Fiu], Maruchedda 2 [Ma2] and Munachedda Nera [MuN]) were discriminated with percentage lower than 80%. These results prove that, also for this class of beans, the genetic diversity is also clearly expressed in the phenotype and that the measured morpho-colorimetric features are objectively discriminant. As expected, in addition to the mean seed weight that shown the highest *F-to-remove* value, 17 of the first 20 parameters, chosen by the identification system to implement the discriminant functions used to distinguish the bi-colored coat bean accessions, are related to the seed coat color and texture. The bi-colored bean landraces analyzed in this study are characterized by a very wide chromatic seed coat variability, also visually identifiable as reported in supplementary information (Suppl. Info. 1). Although landraces have to be considered as mixtures of genotypes, sometimes the phenotypic expression does not fully reflect the intrinsic genetic differences (Harlan, 1975; Hawkes, 1983; Payne et al., 1984; Martin and Adams, 1987; Rieger et al., 1991; Prospéri et al., 1994). Other times, the genetic variability within the same landrace is too low, in comparison to that between two landraces. This results in the possibility to identify phenotypical characters, such as seed color, that allow a clear discrimination also among landraces.

Discrimination for geographical areas of provenance

In order to evaluate the possibility to identify the bean landraces origin, each of the three main color categories of beans (white, mono-colured and bi-colored seeds) were split

Table 1. Percentage of correct identification among white (above) and mono-colored (below) coat bean landraces. Landraces cropped in different localities and/or different years, were considered as same landrace. In parenthesis, the number of analysed seeds.

	White coat bean landraces													
	Bia	BiP	Can	CaP	CaG	Fag	PhC	PiP	Pur	RiB	RiG	Ton	Tri	Total
Bia	95.2 (809)	-	-	-	1.5 (13)	2.8 (24)	-	-	0.4 (3)	-	0.1 (1)	-	-	100.0 (850)
BiP	-	89.6 (95)	-	-	-	-	-	-	-	5.7 (6)	-	4.7 (5)	-	100.0 (106)
Can	-	-	91.4 (127)	-	8.6 (12)	-	-	-	-	-	-	-	-	100.0 (139)
CaP	1.9 (3)	-	4.9 (8)	85.8 (139)	3.1 (5)	-	-	4.3 (7)	-	-	-	-	-	100.0 (162)
CaG	13.0 (13)	-	-	-	87.0 (87)	-	-	-	-	-	-	-	-	100.0 (100)
Fag	0.1 (1)	-	-	-	-	98.2 (1106)	-	-	1.6 (18)	-	-	-	0.1 (1)	100.0 1126)
PhC	-	-	-	-	-	-	95.2 (20)	-	-	4.8 (1)	-	-	-	100.0 (21)
PiP	1.4 (3)	-	-	7.2 (16)	1.8 (4)	-	-	88.2 (195)	-	-	0.9 (2)	-	-	100.0 (221)
Pur	-	-	-	-	0.2 (2)	2.1 (29)	-	-	97.7 (1340)	-	-	-	-	100.0 (1371)
RiB	-	7.2 (12)	-	-	-	-	-	-	-	90.2 (119)	-	0.8 (1)	-	100.0 (132)
RiG	-	-	-	-	-	-	-	-	-	-	93.2 (109)	6.8 (8)	-	100.0 (117)
Ton	-	0.3 (3)	-	-	-	-	-	-	-	0.1 (1)	1.6 (17)	98.1 (1079)	-	100.0 (1100)
Tri	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (113)	100.0 (113)
Overall														96.0 (5558)

	Mono-colored coat bean landraces								
	Cr1	Gia	Moi	SaR	Tab	Vel	Ver	Zol	Total
Cr1	97.6 (41)	-	-	2.4 (1)	-	-	-	-	100.0 (42)
Gia	-	100.0 (675)	-	-	-	-	-	-	100.0 (675)
Moi	-	-	100.0 (428)	-	-	-	-	-	100.0 (428)
SaR	1.7 (3)	-	-	98.3 (172)	-	-	-	-	100.0 (175)
Tab	-	-	-	-	100.0 (135)	-	-	-	100.0 (135)
Vel	-	-	-	-	-	100.0 (294)	-	-	100.0 (294)
Ver	-	-	-	-	-	-	99.4 (172)	0.6 (1)	100.0 (173)
Zol	-	-	-	-	-	-	1.6 (1)	98.4 (60)	100.0 (61)
Overall									99.7 (1983)

Table 2. Percentage of correct identification among bi-colored coat bean landraces. Landraces cropped in different localities and/or different years, were considered as same landrace. In parenthesis, the number of analysed seeds.

	BaB	BaN	Bil	Bor	BoB	CaR	Ciu	Cr2	DeC	Fiu	Gio	Lam	Lar	LaQ	Lup	Ma1	Ma2	Mas	Muc	MuN	Mus	Nas	NaN	NaR	NaV	PaN	PaR	RoL	Sal	SaM	Sch	Sci	Scr	Str	Tuv	TuR	Vio	Total
BaB	**95.1 (370)**	-	-	-	4.4 (17)	-	-	-	-	-	-	-	-	-	-	-	-	-	0.3 (1)	-	-	-	-	0.3 (1)	-	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (389)
BaN	-	**94.4 (304)**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.2 (7)	-	0.6 (2)	0.9 (3)	-	-	-	-	1.9 (6)	-	-	-	-	-	-	-	-	-	-	-	100.0 (322)
Bil	-	-	**76.7 (194)**	3.2 (8)	-	-	-	0.4 (1)	-	-	0.8 (2)	14.2 (36)	0.8 (2)	2.0 (5)	-	-	-	-	-	0.8 (2)	-	-	-	-	-	-	-	-	-	-	-	-	-	1.2 (3)	-	-	-	100.0 (253)
Bor	-	-	-	**92.8 (90)**	-	-	-	-	-	-	1.0 (1)	-	2.1 (2)	-	-	-	3.1 (3)	-	-	-	-	-	-	-	-	-	-	-	-	1.0 (1)	-	-	-	-	-	-	-	100.0 (97)
BoB	0.5 (1)	-	3.3 (7)	12.0 (25)	**70.8 (148)**	-	1.0 (2)	-	-	4.3 (9)	-	1.4 (3)	-	-	-	1.4 (3)	-	-	-	-	-	-	-	-	-	-	-	-	1.4 (3)	0.5 (1)	1.4 (3)	-	0.5 (1)	1.4 (3)	-	-	-	100.0 (209)
CaR	-	-	-	-	-	**99.1 (110)**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.9 (1)	-	-	-	-	-	100.0 (111)
Ciu	-	-	-	-	-	-	**99.4 (171)**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.6 (1)	-	-	-	-	100.0 (172)
Cr2	-	-	-	-	-	-	-	**89.3 (25)**	-	-	-	-	-	-	3.6 (1)	-	-	-	-	-	-	-	-	-	-	-	-	7.1 (2)	-	-	-	-	-	-	-	-	-	100.0 (28)
DeC	-	-	-	-	-	-	-	-	**98.1 (106)**	-	-	-	-	-	-	1.9 (2)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (108)
Fiu	-	-	-	-	16.0 (8)	-	-	-	-	**76.0 (38)**	-	-	-	-	-	4.0 (2)	-	-	-	-	-	-	-	-	-	-	-	-	2.0 (1)	2.0 (1)	-	-	-	-	-	-	-	100.0 (50)
Gio	-	-	-	-	2.6 (1)	-	-	-	-	-	**89.5 (34)**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7.9 (3)	-	-	-	-	-	-	-	100.0 (38)
Lam	-	-	11.3 (31)	-	-	-	-	0.4 (1)	-	-	-	**88.0 (242)**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.4 (1)	-	-	-	100.0 (275)
Lar	-	-	-	0.1 (2)	0.3 (6)	0.0 (1)	0.0 (1)	-	0.1 (2)	0.0 (1)	-	-	**97.3 (2141)**	1.0 (22)	-	-	0.8 (18)	-	-	-	-	-	-	-	-	-	-	-	-	-	0.2 (4)	-	-	0.1 (2)	-	-	-	100.0 (2200)
LaQ	-	-	-	-	-	-	-	-	-	-	-	0.1 (1)	0.2 (2)	**99.7 (947)**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (950)
Lup	-	-	-	-	0.8 (3)	-	2.0 (8)	0.5 (2)	-	0.3 (1)	-	-	-	-	**90.5 (354)**	-	-	-	-	-	-	-	-	-	-	-	-	0.3 (1)	2.3 (9)	-	0.5 (2)	-	1.0 (4)	1.5 (6)	-	-	0.3 (1)	100.0 (391)
Ma1	0.5 (1)	-	-	-	-	-	1.6 (3)	-	3.2 (6)	-	-	-	-	-	-	**94.6 (176)**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (186)
Ma2	-	-	-	4.8 (6)	1.6 (2)	-	-	-	-	-	-	-	23.2 (29)	-	-	-	**68.8 (86)**	-	-	-	-	-	-	-	-	-	-	-	-	1.6 (2)	-	-	-	-	-	-	-	100.0 (125)
Mas	-	0.5 (1)	-	-	-	0.9 (2)	-	-	-	-	-	-	-	-	-	-	-	**90.0 (197)**	-	3.7 (8)	0.5 (1)	-	-	-	-	0.5 (1)	-	-	-	-	-	-	-	-	-	4.1 (9)	-	100.0 (219)
Muc	1.7 (3)	-	-	-	2.9 (5)	-	-	-	-	0.6 (1)	-	-	-	-	-	-	-	-	**93.1 (163)**	-	-	-	-	-	-	-	-	-	-	1.1 (2)	-	-	-	-	-	-	-	100.0 (175)
MuN	-	4.7 (4)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	12.9 (11)	-	**78.8 (67)**	-	-	-	-	-	3.5 (3)	-	-	-	-	-	-	-	-	-	-	-	100.0 (85)
Mus	-	-	-	-	-	-	-	-	-	-	-	-	-	0.2 (1)	-	-	-	-	-	-	**83.6 (351)**	2.1 (9)	4.0 (17)	1.2 (5)	3.6 (15)	-	-	-	-	-	0.5 (2)	0.7 (3)	-	-	4.0 (17)	-	-	100.0 (420)
Nas	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	4.1 (2)	**87.8 (43)**	-	8.2 (4)	-	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (49)
NaN	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**100.0 (75)**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (75)
NaR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	7.1 (6)	-	**87.1 (74)**	5.9 (5)	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (85)
NaV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.5 (3)	-	3.4 (7)	**94.1 (193)**	-	-	-	-	-	-	-	-	-	-	1.0 (2)	-	100.0 (205)
PaN	-	6.5 (11)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.9 (5)	-	0.6 (1)	5.9 (10)	-	-	-	1.2 (2)	**82.9 (141)**	-	-	-	-	-	-	-	-	-	-	-	100.0 (170)
PaR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**96.2 (101)**	-	-	-	1.0 (1)	-	-	-	1.0 (1)	1.9 (2)	-	100.0 (105)
RoL	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.3 (2)	-	-	-	-	-	-	-	-	-	-	-	**98.7 (75)**	1.3 (1)	-	-	-	-	-	-	-	-	100.0 (76)
Sal	-	-	-	-	1.1 (1)	-	1.1 (1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**90.8 (79)**	-	-	-	4.6 (4)	-	-	-	-	100.0 (87)
SaM	-	-	-	1.0 (1)	1.0 (1)	-	-	-	-	2.0 (2)	7.0 (7)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**89.0 (89)**	-	-	-	-	-	-	-	100.0 (100)
Sch	-	-	-	-	0.3 (7)	-	0.4 (9)	0.0 (1)	0.0 (1)	0.0 (1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**99.2 (2256)**	-	-	-	-	-	-	100.0 (2275)
Sci	-	-	-	-	-	0.2 (1)	-	-	-	-	-	-	-	-	0.6 (3)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.2 (1)	**98.8 (491)**	0.2 (1)	-	-	-	-	100.0 (497)
Scr	-	-	-	-	-	-	3.1 (3)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2.1 (2)	7.3 (7)	-	2.1 (2)	-	**85.4 (82)**	-	-	-	-	100.0 (96)
Str	-	-	0.4 (1)	0.8 (2)	2.0 (5)	-	0.4 (1)	-	-	0.8 (2)	-	-	1.2 (3)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**94.3 (232)**	-	-	-	100.0 (246)
Tuv	0.5 (1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.5 (1)	1.0 (2)	-	3.3 (7)	-	-	3.8 (8)	-	0.5 (1)	-	-	-	-	-	-	-	-	**90.4 (189)**	-	-	100.0 (209)
TuR	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1.3 (2)	-	1.9 (3)	-	-	-	-	-	-	-	-	**96.9 (155)**	-	100.0 (160)
Vio	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.9 (1)	-	-	0.9 (1)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**98.2 (110)**	100.0 (112)
Overall																																						**94.3 (11350)**

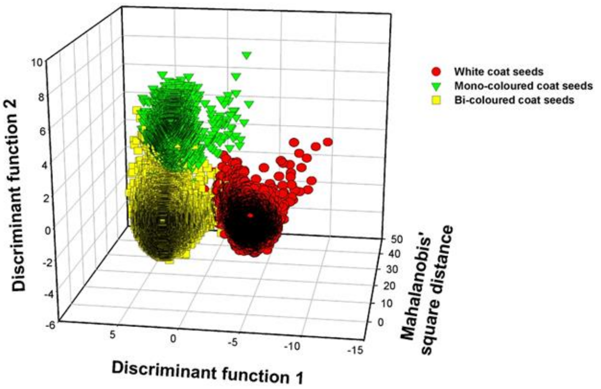

Fig 1. 3D graphical representation of the discriminant analysis for bean landraces distinguished for seed coat main color.

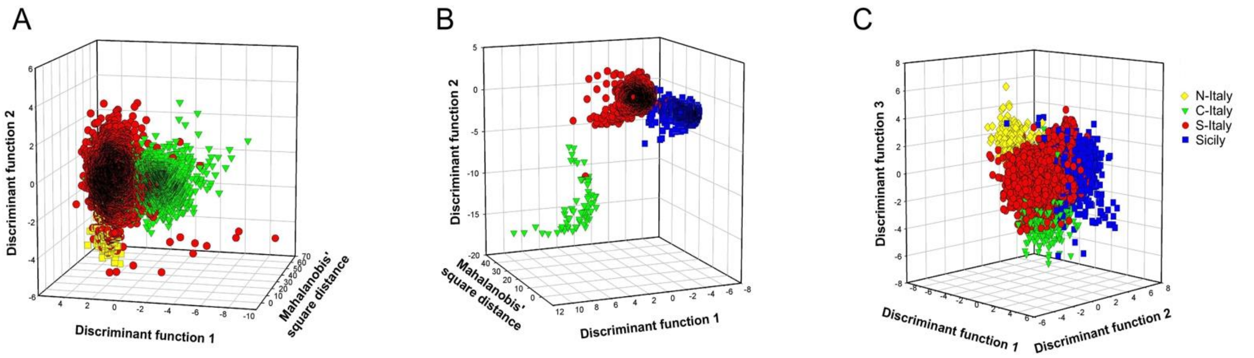

Fig 2. 3D graphical representations of the geographical areas discrimination for the three main color categories: white (A), mono-colored (B) and bi-colored (C) beans coat. No landraces from Sicily were available for white bean coat class, and no landraces from N-Italy were available for bi-colored bean coat class.

in four groups, according to the cropping geographical areas: Northern (N), Central (C), Southern Italy (S) and Sicily, considering this aspect interesting and useful for traceability purpose. It should be highlighted that only the bi-colored coat bean landraces were present into the four geographical areas, whereas the white coat bean landraces were not collected in Sicily and the mono-colored ones were not sampled in N-Italy. Figure 2 reports the graphical representation of the geographical areas discrimination for the three main color categories of bean coat. Overall performances of 92.8%, 98.8% and 88.2% were recorded for white, mono-colored and bi-colored coat beans accessions, respectively (data not shown). Further comparisons were implemented among bean landraces collected in the same geographical area, distinguishing for seed coat color. Regarding the white coat beans, the percentages of correct classification were 99.2% and 99.0% for C-Italy and S-Italy accessions, respectively (data not shown; Fig. 3), recording the minimum value (95.7%) both for Cannellino di Pisa [CaP] (C-Italy) and for Riso Giallo [RiG] (S-Italy). Discrimination among white coat landraces was not possible for N-Italy, because only Bianco di Pigna [BiP] belonged to this geographical group. Both S-Italy and Sicily mono-colored coat bean landraces were perfectly distinguished (data not shown; Fig. 4). Thelandrace Zolfino [Zol] was the only one accession cropped in C-Italy. Classification performance for bi-colored coat beans is shown in table 3. The overall classification of 95.0% was achieved for the N-Italy group, in which the best result was obtained for Saluggia [Sal] and Stregone del Piemonte [Str], whose percentages of correct classification reached 100.0%; while for Billò [Bil] and Lamon [Lam], performances of 93.7% and 90.2% were respectively achieved.

The overall classification for C-Italy beans was 97.8%. The landraces Della Chiesa [DeC], Mascherino [Mas] and Rosso di Lucca [RoL] were perfectly identified, while the remaining landraces reached percentages of correct identification ranged between 91.4% (Borlotto Bianco [BoB]) and 99.0% (Borlotto [Bor] and Scritto di Lucca [Scr]) (Table. 3). The S-Italy group of beans was the largest, with 18 different accessions. They were well identified in 97.8% of cases, recording values higher than 88%, except for Maruchedda [Ma2] that reached 84.8% of correct classification, showing the highest misattribution with Lardariello [Lar] (12.8%) (Table. 3). Finally, classification results about the Sicilian bean landraces are given. The overall correct classification was 99.5%. Badda Bianca [BaB], Fiumara [Fiu], Scicli [Sci] and Viola [Vio] were perfectly identified, while for the other landraces, values included between 89.3% (Crucchittu 2 [Cr2]) and 99.4% (Badda Niura [BaN]) were recorded (Table 3).

The obtained results seem to prove the possibility to identify the bean landraces origin on the basis of morpho-colorimetric features of seeds. These achievements are probably due to the phenotypic expression that not exclusively results from the genotype but also from effect of the growing land where they originated and evolved; as well as climatic conditions and particular agronomical practices, historically applied in some regions and not in others. Landraces of beans, such as of any other crop, consist of seed material phenotypically very susceptible and responsive to biotic and abiotic environmental

Table 3. Percentage of correct identification among bi-colored coat bean landraces distinguished for geographical area. In parenthesis, the number of analysed seeds.

N - Italy	Bil	Lam	Sal	Str	Total
Bil	93.7 (237)	5.9 (15)	-	0.4 (1)	100.0 (253)
Lam	9.8 (27)	90.2 (248)	-	-	100.0 (275)
Sal	-	-	100.0 (87)	-	100.0 (87)
Str	-	-	-	100.0 (246)	100.0 (246)
Overall					95.0 (861)

C - Italy	Bor	BoB	DeC	Lup	Mas	RoL	Scr	Total
Bor	99.0 (96)	1.0 (1)	-	-	-	-	-	100.0 (97)
BoB	6.7 (14)	91.4 (191)	0.5 (1)	1.4 (3)	-	-	-	100.0 (209)
DeC	-	-	100.0 (108)	-	-	-	-	100.0 (108)
Lup	-	0.5 (2)	-	98.5 (385)	-	-	1.0 (4)	100.0 (391)
Mas	-	-	-	-	100.0 (219)	-	-	100.0 (219)
RoL	-	-	-	-	-	100.0 (76)	-	100.0 (76)
Scr	-	-	-	-	-	1.0 (1)	99.0 (95)	100.0 (96)
Overall								97.8 (1196)

S - Italy	CaR	Ciu	Lar	LaQ	Ma1	Ma2	Muc	MuN	Nas	NaN	NaR	NaV	PaN	PaR	SaM	Sch	Tuv	TuR	Total
CaR	99.1 (110)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.9 (1)	-	100.0 (111)
Ciu	-	99.4 (171)	-	-	0.6 (1)	-	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (172)
Lar	-	0.0 (1)	98.3 (2162)	0.7 (15)	-	0.9 (20)	-	-	-	-	-	-	-	-	-	0.1 (2)	-	-	100.0 (2200)
LaQ	-	-	0.2 (2)	99.8 (948)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (950)
Ma1	-	2.7 (5)	-	-	97.3 (181)	84.8 (106)	-	-	-	-	-	-	-	-	-	-	-	-	100.0 (186)
Ma2	-	-	12.8 (16)	-	-	-	-	-	-	-	-	-	-	-	2.4 (3)	-	-	-	100.0 (125)
Muc	-	0.6 (1)	-	-	1.1 (2)	-	96.0 (168)	-	-	-	-	-	-	-	1.7 (3)	-	0.6 (1)	-	100.0 (175)
MuN	1.2 (1)	-	-	-	-	-	-	89.4 (76)	-	-	-	-	8.2 (7)	-	-	-	-	1.2 (1)	100.0 (85)
Nas	2.0 (1)	-	-	-	-	-	-	-	89.8 (44)	-	8.2 (4)	-	-	-	-	-	-	-	100.0 (49)
NaN	-	-	-	-	-	-	-	-	-	100.0 (75)	-	-	-	-	-	-	-	-	100.0 (75)
NaR	-	-	-	-	-	-	-	-	8.2 (7)	-	88.2 (75)	3.5 (3)	-	-	-	-	-	-	100.0 (85)
NaV	-	-	-	-	-	-	-	-	2.0 (4)	-	2.0 (4)	95.6 (196)	-	-	-	-	-	0.5 (1)	100.0 (205)
PaN	-	-	-	-	-	-	-	1.2 (2)	-	-	0.6 (1)	4.1 (7)	94.1 (160)	-	-	-	-	-	100.0 (170)
PaR	-	-	-	-	-	-	-	-	-	-	-	-	-	92.4 (97)	-	1.0 (1)	1.0 (1)	5.7 (6)	100.0 (105)
SaM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	99.0 (99)	-	-	-	100.0 (100)
Sch	-	0.4	-	-	-	-	-	-	-	-	-	-	-	-	-	99.6	-	0.0	100.0 (2275)

		(8)														(2266)		(1)	
Tuv	-	-	-	-	-	1.4 (3)	1.4 (3)	-	-	1.0 (2)	4.3 (9)	-	1.0 (2)	-	-	-	92.3 (193)	-	100.0 (209)
TuR	-	-	-	-	-	0.6 (1)	0.6 (1)	-	-	-	-	2.5 (4)	-	1.9 (3)	-	-	1.3 (2)	93.8 (150)	100.0 (160)
Overall																			97.8 (7437)

Sicily	BaB	BaN	Cr2	Fiu	Gio	Mus	Sci	Vio	Total
BaB	100.0 (389)	-	-	-	-	-	-	-	100.0 (389)
BaN	-	99.4 (320)	-	-	-	0.6 (2)	-	-	100.0 (322)
Cr2	-	-	89.3 (25)	3.6 (1)	3.6 (1)	-	-	3.6 (1)	100.0 (28)
Fiu	-	-	-	100.0 (50)	-	-	-	-	100.0 (50)
Gio	-	-	-	5.3 (2)	94.7 (36)	-	-	-	100.0 (38)
Mus	0.5 (2)	-	-	-	-	99.3 (417)	0.2 (1)	-	100.0 (420)
Sci	-	-	-	-	-	-	100.0 (497)	-	100.0 (497)
Vio	-	-	-	-	-	-	-	100.0 (112)	100.0 (112)
Overall									99.5 (1856)

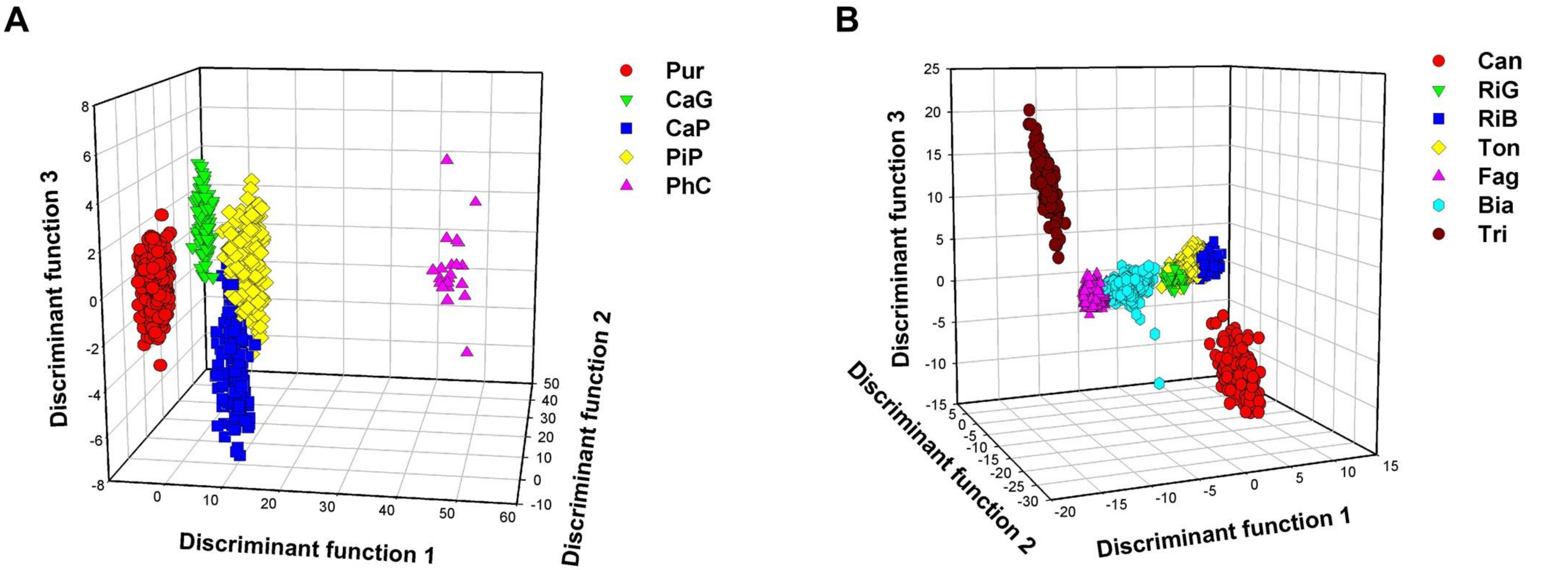

Fig 3. 3D graphical representation of the discriminant analysis among white coat beans landraces collected in C-Italy (A) and S-Italy (B).

Table 4. Percentage of correct identification among different harvest years, for the landraces Borlotto Bianco [BoB], Moitano [Moi] and Lupinaro [Lup]. In parenthesis, the number of analysed seeds.

	BoB 2004	BoB 2005		Total
BoB 2004	100.0 (32)	-		100.0 (32)
BoB 2005	-	100.0 (177)		100.0 (177)
Overall				100.0 (209)
	Moi 2007	Moi 2008		Total
Moi 2007	100.0 (365)	-		100.0 (365)
Moi 2008	-	100.0 (63)		100.0 (63)
Overall				100.0 (428)
	Lup 2004	Lup 2005	Lup 2006	Total
Lup 2004	96.0 (24)	-	4.0 (1)	100.0 (25)
Lup 2005	-	86.3 (151)	13.7 (24)	100.0 (175)
Lup 2006	0.5 (1)	7.3 (14)	92.2 (176)	100.0 (191)
Overall				89.8 (391)

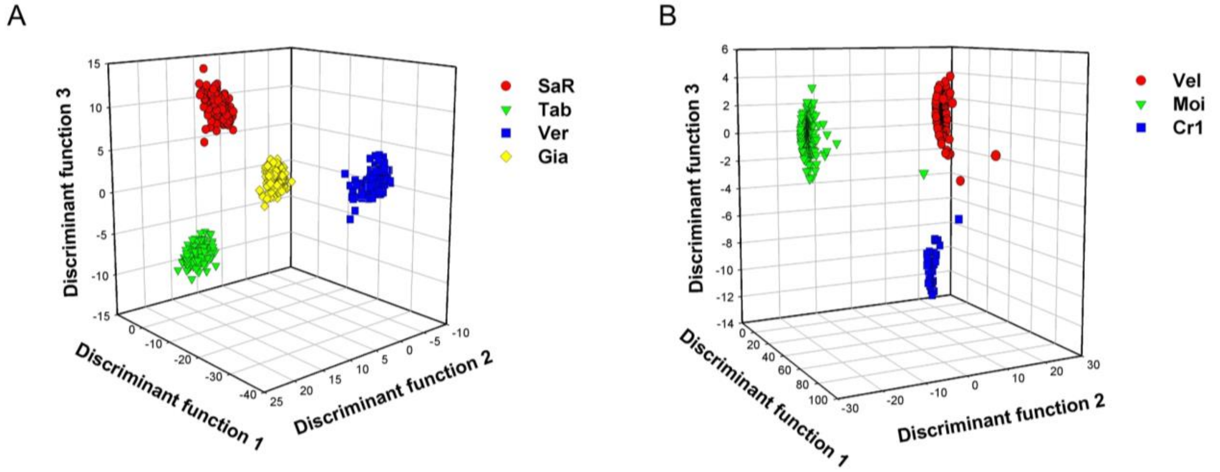

Fig 4. 3D graphical representation of the discriminant analysis among mono-colored coat beans landraces collected in S-Italy (A) and Sicily (B).

factors (Karaköy et al., 2014; Scarano et al., 2014). Moreover, it is known that the phenotype of an organism is dynamic and conditional, representing a complex set of responses to a multi-dimensional pattern of endogenous and exogenous signals that are integrated over the evolutionary and developmental life history of an individual. Phenotypic information can be envisioned as a continuous stream of data that changes over the course of development of species, populations or individuals in response to different environmental conditions (Cobb et al., 2013). For this reason, the work moved to verify the possibility to differentiate some accessions according to their harvest years, cultivation regions and applied agronomic techniques.

Interaction Phenotype × Environment

The effect of the cropping year

According to the collected data, comparisons between two cultivation years, 2004/2005 and 2007/2008, were developed for Borlotto Bianco [BoB] and Moitano [Moi] respectively, and a further one, among three cultivation years, 2004/2005/2006, for Lupinaro [Lup] (Table 4). As shown, the statistical classifiers implemented for Borlotto Bianco [BoB] and Moitano [Moi] landraces, allowed perfect identifications; while an overall percentage of correct classification of 89.8% was recorded for Lupinaro [Lup], investigated for three consecutive years.

These results confirm the hypothesis that the cropping year and the relative climatic conditions affect the phenotypic expression of the seeds, although specific identifying characters are preserved.

The effect of the cultivation region

Table 5 reports the classification performance among the two cultivation regions for the landraces Fiumara [Fiu], Mascherino [Mas], Mussuniuru [Mus] and Purgatorio [Pur]. Perfect identification performances were reached for Fiumara [Fiu], Mascherino [Mas] and Mussuniuru [Mus]. Also the comparison between Purgatorio seeds from Umbria and Lazio allowed to achieve a very high performance (99.9%), misclassifying only one seed over the 1371. As reported in supplementary information (Suppl. Info. 1), Mascherino [Mas] and Purgatorio [Pur] landraces were collected in different provinces of Central Italy, then it is plausible to suppose that the geographical distance between the localities could explain the clear found differentiation. On the other hand, Fiumara [Fiu] and Mussuniuru [Mus] were cropped in the same territory, in close areas with comparable pedo-climatic conditions. In this case, being the seed weight the most powerful feature, the perfect discrimination between the

Table 5. Percentage of correct identification between bean landraces harvested in different localities. In parenthesis, the number of analysed seeds.

Fiu	S. Pietro Patti (ME) Sicily	Raccuia (ME) Sicily	Total
S. Pietro Patti (ME) - Sicily	100.0 (24)	-	100.0 (24)
Raccuia (ME) - Sicily	-	100.0 (26)	100.0 (26)
Overall			100.0 (50)

Mas	Pisa Tuscany	Garfagnana (LU) Tuscany	Total
Pisa (PI) - Tuscany	100.0 (110)	-	100.0 (110)
Garfagnana (LU) - Tuscany	-	100.0 (109)	100.0 (109)
Overall			100.0 (219)

Mus	S. Pietro Patti (ME) Sicily	Raccuia (ME)) Sicily	Total
S. Pietro Patti (ME) - Sicily	100.0 (68)	-	100.0 (68)
Raccuia Sinagna (ME) - Sicily	-	100.0 (352)	100.0 (352)
Overall			100.0 (420)

Pur	Colfiorito di Foligno (PG) Umbria	Gradoli (VT) Lazio	Total
Colfiorito di Foligno (PG) - Umbria	99.8 (545)	0.2 (1)	100.0 (546)
Gradoli (VT) - Lazio	-	100.0 (825)	100.0 (825)
Overall			99.9 (1371)

Fig 5. Geographical distribution of the sampling sites.

Table 6. Percentage of correct identification for the landrace Schiucchiuraliedd [Sch] cropped in the same locality applying different agronomical practices (through row seeding in organic fertilization or by pocket drilling without fertilization and irrigation; below). In parenthesis, the number of analysed seeds.

	Row seeding in organic fertilization	Pocket drilling without fertilization and irrigation	Total
Row seeding in organic fertilization	100.0 (1375)	-	100.0 (1375)
Pocket drilling without fertilization and irrigation	-	100.0 (900)	100.0 (900)
Overall			100.0 (1275)

seed lots cropped in the two different areas could be due to different agronomical treatments applied to the two crops.
Comparing the above data with those obtained by Venora et al. (2009b), no significant difference can be detected in the classifier performance used to discriminate bean accessions according to the cropping year. Differently, matching up to the results achieved from the comparison between the landrace Purgatorio grown in Umbria and the same landrace cropped in Lazio, it is possible to note the effect of the added parameters for the seeds discriminant analysis. The image analysis macro used by Venora et al. (2009b) allowed to correctly identify 68.1% of the landrace Purgatorio from Umbria, misclassifying as Purgatorio from Lazio 31.9% of the cases. The increasing of the analyzed seeds number, together with the improvements made to the macro, adding the mean seed weight, the 78 EFDs and 22 Haralick's descriptors, released as result 99.9% of correct classification, proving an higher ability of classifier.

The effect of the agricultural practices

A further comparison was carried out to verify the effect of different agricultural practices in the seed phenotyping, assessing the possibility to discriminate between seed lots of a same landrace, grown in the same locality applying different agricultural practices. The landrace Schiucchiuraliedd [Sch] was cropped in the same locality through row seeding in organic fertilization or by pocket drilling without fertilization and irrigation (Table 6). In this case, an overall correct identification percentage of 100.0% of correct recognition was achieved. One more time, even though some phenotypic peculiarities of the seed remain unchanged, the reached results prove the great implications that, both environment and agronomic treatments, have on the seed morpho-colorimetric characters. As mixture of genotypes, landraces are distinct but variable populations, characterized by a specific adaptation to the environmental conditions of the cultivation area (tolerant to the biotic and abiotic stresses of that area). They are closely associated with the uses, knowledge, habits, dialects, and celebrations of the people who developed and continue to grow it, also applying different agricultural practices (Negri et al., 2009; Polegri and Negri, 2010).

Materials and Methods

Seed samples collection and acquisition

Bean samples of 58 Italian landraces were investigated and characterized in this study. Four of these landraces (Fiumara, Mascherino, Mussuniuru and Purgatorio) were collected from two cultivation regions, in order to evaluate the effect of the geographical position, while the landraces Borlotto Bianco, Moitano and Lupinaro were monitored and collected for two or three consecutive harvest years (2004/2005; 2007/2008; 2004/2005/2006), in order to investigate environmental or

Fig 6. Representative bean samples of some of the landraces considered in the study.

seasonal differences. In addition, the landrace Schiucchiuraliedd was analyzed as it was cultivated applying different agricultural techniques, in the same geographical locality: row seeding in organic fertilization and pocket drilling without fertilization and irrigation management. A total of 67 accessions were investigated. The studied bean samples and their main seed characteristics are given as supplementary information (Supplemetary Table 1). Figure 5 reports the geographical position of the sampling sites and in figure 6, some of the studied bean landraces are shown.
Digital images of beans were acquired, using a flatbed scanner (ScanMaker 9800 XL, Microtek Denver, CO), following the procedure reported in Pinna et al. (2014) and

processed using the software package KS-400 V. 3.0 (Carl Zeiss, Vision, Oberkochen, Germany). A total of 18,893 bean seeds were analyzed.

Elliptic Fourier (EFDs) and Haralick's descriptors

In order to increase the discrimination power, the macro specifically developed by Venora et al., (2009b) for the characterization of bean seeds,was further enhanced adding algorithms that allow to compute the Elliptic Fourier Descriptors (EFDs) for each analyzed seed, obtaining further 78 quantitative variables. As described by Orrù et al. (2012), this method allows to define the boundary of the seed projection, as an array of complex numbers which correspond to the pixels position of the seed boundary. According many authors about the use of number of harmonics for an optimal description of seed outlines, 20 harmonics were used to define the seed boundaries (Orrù et al., 2013). Moreover, the macro was improved including algorithms able to calculate 11 Haralick's descriptors with the relative standard deviation values for each seed. These parameters are generally used when the objects in the images cannot be separated due to indefinite grey values variations. In these cases, the evaluation of texture, tone and context allows to define the spatial distribution of the image intensities and discrete tonal features. When a small area of the image has little variation of discrete tonal features, the dominant property of that area is grey tone. When a small area has wide variation of discrete tonal features, the dominant property of that area is texture (Haralick and Shapiro, 1991). According Haralick et al. (1973), the concept of tone is based on varying shades of grey of resolution cells in a photographic image, while texture is concerned with the spatial (statistical) distribution of grey tones. Texture and tone are not independent concepts; rather, they bear an inextricable relationship to one another very much like the relationship between a particle and a wave. Context, texture and tone are always present in the image, although at times one property can dominate the others. The basis for these features is the gray-level co-occurrence matrix (*G* in equation 1). This matrix is square with dimension N_g, where N_g is the number of gray 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 to be the probability that a pixel with value *i* will be found adjacent to a pixel of value *j*.

$$G = \begin{bmatrix} p(1,1) & p(1,2) & \dots & p(1,N_g) \\ p(2,1) & p(2,2) & \cdots & p(2,N_g) \\ \vdots & \vdots & \ddots & \vdots \\ p(N_g,1) & p(N_g,2) & \dots & p,(N_g,N_g) \end{bmatrix} \qquad \text{(eq. 1)}$$

The 11 Haralick's descriptors measured on each seed to mathematically describe the surface texture and all the other morpho-colorimetric characters are available as supplementary information (Supplementary Table 2 and 3).
Mean seed weight of each seed lot was also included to increase the discriminant power of the statistical analysis. It was determined weighing 20 seeds for ten times, on a four decimal places scale.

Statistical analysis

The achieved data were used to build a global database, including morpho-colorimetric, EFDs and Haralick's descriptors and mean seed weight. Statistical elaborations were executed using SPSS software package release 15 (SPSS, 2007), applying the same stepwise Linear Discriminant Analysis (LDA) algorithm suggested by Grillo et al. (2012) to identify and discriminate among the investigated bean 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., 2009; Venora et al., 2009b; Holden et al., 2011; Rencher and Christensen, 2012; Kuhn and Johnson, 2013). 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 (Venora et al., 2007; 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. This procedure, also called rotation estimation (Picard and Cook, 1984; Kohavi, 1995), was applied, both to evaluate the performance and to validate any classifier. The validation procedure here used is the Leave-One-Out Cross-Validation (LOOCV). 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). To graphically highlight the differences among seed 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 Mahalanobis' distance (Mahalanobis, 1936) were used (Bacchetta et al., 2008).

Conclusions

The achievements allow demonstrating the usefulness of the discrimination system based on seed phenotypic characters, for the identification and classification of bean accessions. The technique here proposed, conveniently sustained by a conspicuous database, can be undoubtedly considered a helpful tool as a support for any other recognized identification systems such as DNA fingerprinting and bar-coding. 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 bean 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. Therefore, the implementation of food traceability mechanisms is particularly relevant for developing countries who wish to increase extending their share in international food trade.

Considering the heterogeneous nature of the seed samples used in this study, in order to validate these preliminary achievements, further trials will have to be conducted focusing on few selected landraces, cropped in many different localities, for many consecutive years and applying different defined agronomical practices.

Acknowledgements

The authors thank Giuseppina Logozzo of Università degli Studi della Basilicata, Scuola di Scienze Agrarie, Potenza (Italy), for kindly providing the bean populations of Basilicata region, South Italy.

References

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

Cobb JN, De Clerck G, Greenberg A, Clark R, McCouch S (2013) Next-generation phenotyping: requirements and strategies for enhancing our understanding of genotype-phenotype relationships and its relevance to crop improvement. Theor Appl Genet. 126:867-887.

De La Fuente M, López-Pedrouso M, Alonso J, Santalla M, De Ron AM, Álvarez G, Zapata C (2012) In-depth characterization of the phaseolin protein diversity of common bean (*Phaseolus vulgaris* L.) based on two-dimensional electrophoresis and mass spectrometry. Food Technol Biotech. 50: 315-325.

Diamond J, Anderson NH, Bartels PH, Montironi R, Hamilton PW (2004) The use of morphological characteristics and texture analysis in the identification of tissue composition in prostatic neoplasia. Hum Pathol. 35:1121-1131.

Diniz AL, Zucchi MI, Santini L, Benchimol-Reis LL, Fungaro MHP, Carneiro Vieira ML (2014) Nucleotide diversity based on phaseolin and iron reductase genes in common bean accessions of different geographical origins. Genome. 57: 69-77.

Dreher K, Khairallah M, Ribaut JM, Morris M (2003) Money matters (I): costs of field and laboratory procedures associated with conventional and marker-assisted maize breeding at CYMMIT. Mol Breeding. 11:221-234.

Gerger A, Smolle J (2003) Diagnostic imaging of melanocytic skin tumors. J Cutan Pathol. 30:247-252.

Granitto PM, Garralda PA, Verdes PF, Ceccato HA (2003) Boosting classifiers for weed seeds identification. J Cereal Sci. 3:34-39.

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

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

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

Grillo O, Miceli C, Venora G (2011) Image analysis tool for vetch varieties identification by seeds inspection. Seed Sci Technol. 39:90-500.

Grisi MC, Blair MW, Gepts P, Brondani C, Pereira PA, Brondani RP (2007) Genetic mapping of a new set of microsatellite markers in a reference common bean (*Phaseolus vulgaris*) population BAT93 x Jalo EEP558. Genet Mol Res. 30:691-706.

Haralick RM, Shanmugam K, Dinstein I (1973) Textural features for image classification. IEEE T Syst Man Cyb. 6:610-621.

Haralick RM, Shapiro LG (1991) Glossary of computer vision terms. Pattern Recogn. 24:69-93.

Harlan JR (1975) Our vanishing genetic resources. Science. 188:618–621.

Hâruta O (2011) Elliptic Fourier analysis of crown shapes in *Quercus petraea* trees. Ann For Res. 54:99-117.

Hastie T, Tibshirani R, Friedman J (2009) Linear methods for classification. In: Hastie T, Tibshirani R, Friedman J (eds) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer, New York.

Hawkes JG (1983) The diversity of crop plants. Harvard Univ. Press ,Cambridge, Mass. (USA).

Holden JE, Finch WH, Kelly K (2011) A comparison of two-group classification methods. Educ Psychol Meas. 715:870-901.

Houle D, Govindaraju DR, Omholt S (2010) Phenomics: the next challenge. Nat Rev Genet. 11:855-866.

Iwata H, Nesumi H, Ninomiya S, Takano Y, Ukai Y (2002) Diallel analysis of leaf shape variations of citrus varieties based on elliptic Fourier descriptors. Breeding Sci. 52:89-94.

Iwata H, Niikura S, Seiji M, Takano Y, Ukai Y (2004) Genetic control of root shape at different growth stages in radish (*Raphanus sativus* L.). Breeding Sci. 54:117-124.

Iwata H, Ebana K, Uga Y,

Hayashi T, Jannink J-L (2010). Genome-wide association study of grain shape variation among Oryza sativa L. germplasms based on elliptic Fourier analysis. Mol Breeding. 25:203–215.

Karaköy T, Baloch FS, Toklu F, Özkan H (2014) Variation for selected morphological and

quality-related traits among 178 faba bean landraces collected from Turkey. Plant Gen Res. 12:5-13.

Kawabata S, Yokoo M, Nii K (2009) Quantitative analysis of corolla shapes and petal contours in single-flower cultivars of *Lisianthus*. Sci Hortic-Amsterdam. 121:206-212.

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Martin TP, Ralescu AL (eds) IJCAI'95: fuzzy logic in artificial intelligence, towards intelligent systems. 14th International Joint conference on artificial intelligence, Montréal Québec, Canada, August 1995. (Lecture notes computer science). Morgan Kaufmann, San Francisco, p 1137.

Kuhn M, Johnson K (2013) Discriminant analysis and other linear classification models. In: Kuhn M, Johnson K (eds) Applied predictive modeling, Springer, New York.

Lioi L, Piergiovanni AR, Pignone D, Puglisi S, Santantonio M, Sonnante G (2005) Genetic diversity of some surviving *on-farm* Italian common bean (*Phaseolus vulgaris* L) landraces. Plant Breeding. 124:576-581.

Lo Bianco M, Ferrer-Gallego P, Grillo O, Laguna E, Venora G, Bacchetta G (2015) Seed image analysis provides evidence of taxonomical differentiation within the Medicago L. sect. Dendrotelis (Fabaceae). Systematics & Biodiversity, 13(5): 484-495.

Logozzo G, Donnoli R, Macaluso L, Papa R, Knüpffer H, Spagnoletti Zeuli P (2006) Analysis of the contribution of Mesoamerican and Andean gene pools to European common bean (*Phaseolus vulgaris* L.) germplasm and strategies to establish a core collection. Genet Resour Crop Ev. 54:1763-1779.

Lootens, Van Waes, Carlier, 2007. Description of the morphology of roots of Chicorium intybus L. partim by means of image analysis: Comparison of Elliptic Fourier Descriptors and classical parameters. Computers and Electronics in Agriculture 58: 164–173.

Mahajan S, Das A, Sardana HK (2015) Image acquisition techniques for assessment of legume quality. Trends Food Sci Tech. 42:116-133.

Mahalanobis PC (1936) On the generalized distance in statistics. Proceedingsof the national Academy of Sciences, India 12:49-55.

Marotti I, Bonetti A, Minelli M, Catizone P, Dinelli G (2007) Characterization of some Italian common bean (*Phaseolus vulgaris* L.) landraces by RAPD, semi-random and ISSR molecular markers. Genet Resour Crop Ev. 54:175-188.

Martin G, Adams MW (1987) Landraces of *Phaseolus vulgaris* in Northern Malawi, I: regional variation. Econ Bot. 41:190-203.

Mercati F, Leone M, Lupini A, Sorgonà A, Bacchi M, Abenavoli MR, Sunseri F (2013) Genetic diversity and population structure of a common bean (*Phaseolus vulgaris* L.) collection from Calabria (Italy). Genet Resour Crop Ev. 60:839-852.

Nanni L, Shi JY, Brahnam S, Lumini A (2010) Protein classification using texture descriptors extracted from the protein backbone image. J Theor Biol. 264:1024-1032.

Negri V (2003) Landraces in central Italy: where and why they are conserved and perspective for their *on-farm* conservation. Genet Resour Crop Ev. 50:871-885.

Negri V, Maxted N, Vetelainen M (2009) European landrace conservation: an introduction. In: Vetelainen M, Negri V, Maxted N (eds) Technical Bullettin n. X. European landraces: *On-farm* conservation, management and use. Bioversity International. Rome, Italy

Orrù M, Grillo O, Venora G, Bacchetta G (2012) Computer vision as a complementary to molecular analysis: grapevines cultivars case study. C R Biol. 335:602-615.

Orrù M, Grillo O, Lovicu G, Venora G, Bacchetta G (2013) Morphological characterisation of *Vitis vinifera* L. seeds by image analysis and comparison with archaeological remains. Veg Hist Archaeobot 22: 231-242.

Park YJ, Lee JK, Kim NS (2009) Simple Sequence Repeat Polymorphisms (SSRPs) for evaluation of molecular diversity and germplasm classification of minor crops. Molecules. 14:4546-4569.

Payne PI, Holt LM, Jackson EA, Law CN, Damania AB (1984) Wheat storage proteins: their genetics and their potential for manipulation by plant breeding. Philos Trans R Soc Lon B Biol Sci. 304:359-371.

Picard R, Cook D (1984) Cross-validation of regression models. J Am Statist Assoc. 79:575-583.

Piergiovanni AR, Cerbino D, Brandi M (2000) The common bean populations from Basilicata (Southern Italy). An evaluation of their variation. Genet Resour Crop Ev. 47:489-495.

Piergiovanni AR, Laghetti G (1999) The common bean landraces from Basilicata (Southern Italy): an example of integrated approach applied to genetic resources management. Genet Resour Crop Ev. 46:47-52.

Piergiovanni AR, Lioi L (2010) Italian common bean landraces: history, genetic diversity and seed quality. Diversity. 2: 837-862.

Polegri L, Negri V (2010) Molecular markers for promoting agro-biodiversity conservation: a case study from Italy. How cowpea landraces were saved from extinction. Genet Resour Crop Ev. 57:867-880.

Prospéri JM, Demarquet F, Angevain M, Mansat P (1994) Evaluation agronomique de variétés de pays de sainfoin (*Onobrychis sativa* L.) originaires du sud-est de la France. Agronomie. 14:285-298.

Reggi L, Tiranti B, Negri V (2013) Italian common bean landraces: diversity and population structure. Genet Resour Crop Ev. 60:1515-1530.

Rencher AC, Christensen WF (2012) Methods of multivariate analysis. 3rd edn, John Wiley & Sons, New York.

Rieger R, Michaelis A, Green MM (1991) Glossary of genetics Classical and molecular. 5th edn. Springer-Verlag, Berlin.

Rodiño AP, Monteagudo AB, De Ron AM, Santalla M (2009) Ancestral landraces of common bean from the south of europe and their agronomical value for breeding programs. Crop Sci. 49:2087-2099.

Scarano D, Rubio F, Ruiz JJ, Rao R, Corrado G (2014) Morphological and genetic diversity among and within common bean (*Phaseolus vulgaris* L.) landraces from the Campania region (Southern Italy). Sci Hortic-Amsterdam. 180:72-78.

Shahin MA, Symons SJ (2003) Lentil type identification using machine vision. Can Biosyst Eng. 45:3.5-3.10.

Sicard D, Nanni L, Porfiri O, Bulfon D, Papa R (2005) Genetic diversity of *Phaseolus vulgaris* L. and *P. coccineus* L. landraces in central Itlay. Plant Breeding. 124:464-472.

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 Crop Prod. 47:232-238.

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.

Venora G, Grillo O, Ravalli C, Cremonini R (2007) Tuscany beans landraces, on-line identifications from seeds inspection by image analysis and Linear Discriminant Analysis. Agrochimica. 51:254-268.

Venora G, Grillo O, Saccone R (2009a) Durum wheat storage centers of Sicily: evaluation of vitreous, starchy and shrunken kernels by image analysis system. J Cereal Sci. 49:429-440.

Venora G, Grillo O, Ravalli C, Cremonini R (2009b) Identification of Italian landraces of beans (*Phaseolus vulgaris* L.), using an image analysis system. Sci Hortic-Amsterdam. 121:410-418.

Yoshioka Y, Honjo M, Iwata H, NINOMIYA S, Ohsawa R (2007) Pattern of geographical variation in petal shape in wild populations of Primula sieboldii E. Morren. Plant Spec Biol. 22: 87–93.