Received: 31 October 2022 | Accepted: 28 June 2023 | Published online: 3 August 2023

DOI: 10.1002/agj2.21423

Agronomy Journal

ORIGINAL ARTICLE

Biometry, Modeling, and Statistics

Variability of morphological descriptors in Sicilian oat (*Avena sativa* L.) populations

Alessandra Carrubba | **Francesca Di Salvo** | **Roberto Marceddu** | **Mauro Sarno**

Department of Agricultural, Food and Forest Sciences, Università degli Studi di Palermo, Palermo, Italy

Correspondence
Alessandra Carrubba, Department of Agricultural, Food and Forest Sciences, Università degli Studi di Palermo, Viale delle Scienze, edif. 4, ingr. L Palermo, PA, 90128, Italy.
Email: alessandra.carrubba@unipa.it

Assigned to Associate Editor Mehdi Rahimi.

Abstract

Due to its rusticity and feasibility of use, oat (*Avena sativa* L.) represents a crucial agronomic and economic resource for many semiarid environments. Presently, the recourse to new commercial varieties has caused a dramatic lowering of areas covered with the traditional local genotypes, and a severe risk of genetic erosion is emerging. To deepen the knowledge about the autochthonous oat populations, an activity of collection and cataloging across semiarid cropping areas was carried out. Sixteen oat populations were collected from different areas of Sicily and put in a field study for two consecutive years (2014 and 2015) in the experimental farm "Sparacia" (Cammarata, Italy). In both years and all populations, 21 morphological characters, related to different aspects of the whole plant or plant parts, were measured as described in the guidelines Community Plant Variety Office—Office Communautaire des Varietes Vegetales (CPVO-OCVV) (rif. CPVO-TP/020/2). Multivariate analysis (MA) was applied to assess the similarity/dissimilarity level among populations, also evaluating the relative discriminatory importance of each selected plant character. Although a strong variability between years did not allow perfect discrimination among genotypes, an association between oat groups emerged based on their prevalent utilization form. Among categorical characters, measurements on glumes and grain provided the best characterization of the populations in both years.

1 | INTRODUCTION

Oat (*Avena sativa* L.; fam. *Poaceae*) is an annual herbaceous crop largely cultivated throughout the world. According to food and agriculture organization of the United Nations (FAO) estimates (FAOSTAT, 2022), in 2020 it covered an area of 9.8 million hectares worldwide, onto which about 25 million tons of grains were harvested. The primary utilization of oat is as human food resource: its seeds (kernels) are the basic ingredient for many food preparations such as puddings, flakes, or breakfast cereals. Oat flour, although not suitable alone for bread-making due to the absence of gluten, may be used to obtain several bakery products after mixing with wheat flour (Butt et al., 2008; Šmídová & Rysová, 2022). Interest in oat has increased since the mid-1980s, due to many healthy claims addressed to its beneficial effects in the frame of the new "nutraceutical" products (Singh et al., 2013). Oat is reputed to be beneficial for human health due to its high-content of dietary fiber, especially β-glucans (Zhang et al., 2021), minerals, and other nutrients. Nowadays, it is

Abbreviations: CPVO-OCVV, Community Plant Variety Office—Office Communautaire des Varietes Vegetales; MA, multivariate analysis; MANOVA, multivariate analysis of variance; MFA, multifactorial analysis; PCA, principal component analysis.

a valuable ingredient for many cosmetic and pharmaceutical items.

Furthermore, oat is a well-known fodder resource: grains are used to feed horses, cattle, pigs, or poultry, and the whole plant is used for grazing, silage, green forage, or hay production. Hence, due to its great plasticity and suitability to many environments and cropping conditions, oat is a true multifunctional crop, capable of accomplishing many roles on a farm (Francia et al., 2006; Sánchez-Martín et al., 2014; Zwer, 2016).

In Sicily, it is traditionally considered an important resource for farms that use this plant both for grazing and grains, as well as in different mixed utilization forms (grazing until tillering stage, then hay or grain). However, the competition with other more profitable crops and some agronomical flaws of the species, such as its lodging tendency, have dramatically reduced its cultivation, insomuch many locally grown populations are presently at serious risk of depletion. Moreover, the recourse to new commercial varieties, mostly coming from breeding activities in Northern and Central Europe, has caused a dramatic lowering in areas covered with the traditional local genotypes, with a consequent strong risk of genetic erosion.

In Sicily, due to the high diversification of habitats, a conspicuous number of landraces have been assessed for many species, including oat. The importance of landraces has been well assessed in many herbaceous crops dealing with a commercial interest, such as durum wheat (*Triticum turgidum* L. subsp. *durum* Desf.; Roselló et al., 2019; Ruisi et al., 2021), barley (*Hordeum vulgare* L.; Jones et al., 2011), common bean (*Phaseolus vulgaris* L.; Piergiovanni & Lioi, 2010), chickpea (*Cicer arietinum* L.; Kumar et al., 2015), and tomato (*Solanum lycopersicum* L.; Caramante et al., 2021). Landraces are recognized as a crucial source of adaptive genes, and their preservation keeps an increasing importance (Villa et al., 2005; Fiore et al., 2019). Hence, several papers were devoted to the evaluation, characterization, and conservation of genetic material in many species where the presence of landraces is outstanding. Research explored plant variability at family level such as in the *Annonaceae* family (Castañeda-Garzón et al., 2016), at genus level (*Capsicum* spp., Junior e Silva et al., 2013), and at species level such as in quinoa (*Chenopodium quinoa* Willd.; Rojas, 2003), cassava (*Manihot esculenta* Crantz; de Oliveira et al., 2014; Silva et al., 2017), cocoa (*Theobroma cacao* L.; Adewale et al., 2013), rosemary (*Salvia rosmarinus* Spenn.; Carrubba et al., 2020), and common bean (Ligarreto, 2013). In most cases, this research focused on detecting similarities and differences among the studied populations, revealing the presence of duplicates within each collection. However, although the newest "-omics" techniques can be successfully applied and represent the future of the germplasm evaluation (Weckwerth et al., 2020), the classical evaluation approach by means of phenotypic selection still bears a deep importance, also because it is often the only method largely applicable by farmers and breeders.

Core Ideas

- In Sicily, many oat landraces are grown and multiplied by local farmers, mostly for animal feeding.
- There is no certainty about the actual extent of difference/similarity among the locally grown populations.
- We analyzed 21 morphological quantitative-qualitative descriptors on 16 oat populations in a 2-year cultivation.
- We used multifactorial analysis (MFA) to assess the relative importance of morphological traits for full characterization of populations.
- A strong variability showed up between cropping years for most of the examined variables.

Multivariate analysis (MA) could be a very useful statistical tool for this purpose (Hardle & Simar, 2019); according to the scope of analysis, it can summarize the overall variability of the examined collection into two (or a few) variables, and it can classify evaluated individuals into groups at various levels of internal similarity (Le Dien & Pagès, 2003).

The first goal (grouping variables) is achieved by means of principal component analysis (PCA; Jolliffe, 2002). The goal of PCA is to transform a large set of variables into a smaller number of hypothetical variables called principal components, which still retain most of the information in the original set, achieving dimensionality reduction. The core of PCA is a singular value decomposition of the correlation matrix of the quantitative variables and the computation of eigenvalues and eigenvectors. These newly obtained linear combinations can be interpreted as a smaller, easier-to-handle, set of indicators.

The PCA, however, works on quantitative characters. When qualitative (categorical) descriptors are involved, multifactorial analysis (MFA) is adopted (Le Dien & Pagès, 2003). Under the same principles of PCA, MFA allows forming a set of orthogonal axes, called factors or dimensions, where observations and variables can be simultaneously displayed, making it easy to discover the information included in data and frequencies. MFA measures the contribution of the morphological characters in terms of mixed variables and defines the distances between populations. Hence, it displays the proximity between observations or between variables.

The second goal (grouping individuals) is achieved through hierarchical cluster analysis (Everitt et al., 2011). Cluster analysis contributes measuring the distances among the populations and grouping them into clusters, so that populations

TABLE 1 *Avena sativa* 2014 and 2015. Synoptic table of the 16 studied populations, selected characters of provenance farms, and prevailing management techniques for oat.

No.	Code	Provenance	Coordinates	Altitude (m a.s.l.)	Use[a]	Sowing		Fertilization (kg ha^{-1})		Weed control[d]
						Method[b]	Rate (kg ha^{-1})	P_2O_5	N[c]	
6	ALCMO	Alcamo (TP)	37°58′44″ N; 12°58′05″ E	250	g, h	R	150	35	18 + 46	Y
13	CLTVT	Caltavuturo (PA)	37°49′11″ N; 13°53′27″ E	600	g, h	S	180	0	0 + 46	N
1	CSTLC	Castel di Lucio (ME)	37°53′08″ N; 14°18′53″ E	750	g, gr + h	S	150	46	18 + 46	N
8	CFLD1	Cefalà Diana 1 (PA)	37°54′56″ N; 13°27′38″ E	650	g	S	150	38	23 + 46	Y
3	CFLD2	Cefalà Diana 2 (PA)	37°54′56″ N; 13°28′00″ E	650	g, h	S	180	46	18 + 46	Y
15	CNTSS	Contessa Entellina (PA)	37°43′46″ N; 13°10′59″ E	550	g	R	180	39	23 + 0	N
14	CRLNE	Corleone (PA)	37°48′50″ N; 13°17′55″ E	600	g	S	200	69	27 + 46	N
5	GBLLN	Gibellina (TP)	37°49′12″ N; 12°57′16″ E	250	g, gr + h	S	200	46	18 + 46	N
9	MLNZZ	Mulinazzo (AG)	37°34′15″ N; 13°31′37″ E	320	g, gr + g, h	S	180	46	18 + 46	N
7	PALBN	Piana Degli Albanesi (PA)	37°59′37″ N; 13°17′17″ E	700	g, gr + h	S	200	38	23 + 46	N
11	RCCMN	Roccamena (PA)	37°50′24″ N; 13°09′19″ E	550	g, h	R	150	46	18 + 46	Y
16	SMBC1	Sambuca 1 (AG)	37°38′52″ N; 13°06′39″ E	300	g, h	S	200	38	23 + 35	Y
12	SCTLD	San Cataldo (CL)	37°29′00″ N; 13°59′17″ E	550	gr + g, gr + h	R	160	64	25 + 55	Y
4	SSTEF	Santo Stefano Quisquina (AG)	37°35′59″ N; 13°31′10″ E	700	gr + g, gr + h	S	200	46	18 + 34	N
2	VLLFR	Villafrati (PA)	37°54′20″ N; 13°29′21″ E	750	g	S	180	30	15 + 46	N
10	VITA	Vita (TP)	37°52′19″ N; 12°49′14″ E	300	gr + g, gr + h	S	200	0	0 + 69	N

[a]Use: g, grain; h, hay; gr, grazing; gr + g, grazing until tillering, then grain; gr + h, grazing until tillering, then hay.
[b]Sowing method: R, on rows; S, spread.
[c]Nitrogen fertilization: kg ha^{-1} split between sowing time (first digit) and on crop (second digit).
[d]Weed control: Y, yes; N, no.

within one cluster are more similar to each other than to populations in other clusters.

The relevant result is the transformation of mixed data (categorical and continuous) in a set of numerical information and the reduction of their dimensionality at the cost of preserving a smaller amount of variability. The technique also allows assessing which dimensions are more relevant for the interpretation of data structure; proximities, between observations and variables projected in these axes, projected in a lower dimensionality, are interpreted as strong relationships. The continuous variables are categorized into classes of values, and the singular value decomposition of the respective groups is implemented; balance of the influences is also required among the multiple sets of variables simultaneously considered.

To our knowledge, only a few works have been aimed, so far, at surveying and cataloging Sicilian oat landraces (Bono et al., 2007, 2008, 2009). In most cases, they only focused on the grain yield-related characters, paying little attention to the morphological and biological traits of plants.

FIGURE 1 Sites of collection of the 16 *Avena sativa* studied populations.

Unlike many other crops, research on the description and study of intraspecific variability of oat is rather scarce, and mostly limited to the areas where the crop has reached a significant importance, such as Turkey (Dumlupinar et al., 2012), India (Kumar et al., 2023; Wagh et al., 2019), or Pakistan (Ihsan et al., 2021). Furthermore, although MA is quite always recognized as the most useful tool to explore such variability, categorical descriptors are never included, probably because their inclusion makes the whole analysis more challenging.

The main aim of this work was to explore the morphobiological diversity in oat populations of Sicily, by means of MA, especially focusing on:

1. to analyze the degree of difference/similarity among the observed populations;
2. to assess the relative importance of morphobiological descriptors for full characterization of oat populations.

2 | MATERIALS AND METHODS

2.1 | Plant material

A survey was carried out in the early 2000s, with the aim of identifying and cataloging farms interested in cultivating oat landraces. About 30 farms were retrieved, all located in N-W Sicily. Through interviews with the farmers, information was gathered about the provenance of oat populations, pedological and climatic characteristics of the farms, farming systems and management, crop utilization (grain, grazing, hay, silage, or double use), and about the cropping technique adopted in the farm (time and methods of sowing, soil management, fertilization, weed control). A 5 kg seed sample was taken and labeled with the name of the collection site. At the end of the survey, the obtained material, consisting of seeds from 16 oat populations collected in different areas of Sicily (Table 1; Figure 1), was put in a field experiment (re-randomized every year) in two consecutive years (2014 and 2015).

2.2 | Experimental site and climatic details

The collection plots were established in the experimental farm "Sparacia" (Cammarata, AG; 37°38′ N–13°46′ E; 415 m s.l.m.) on a clayey soil classified as Chromic Haploxerert (Soil Survey Staff, 2014), with an average slope of about 10%. The primary chemical and physical soil characteristics are reported in Table 2.

Oat populations were sown on February 17, 2014, and February 16, 2015, on a soil previously cultivated with berseem clover and durum wheat in the first and second trial year, respectively.

In summer, the soil seedbed was prepared through a shallow soil work (25 cm). Before sowing, 92 kg ha^{-1} P_2O_5 and

TABLE 2 Sparacia (Cammarata, AG, Italy), 2014 and 2015. Major chemical and physical characteristics of the soil used for oat (*Avena sativa*) cultivation.

Property	Value
Clay (%)	37.94
Silt (%)	24.43
Sand (%)	37.63
pH	8.12
$CaCO_3$ (total)	14.51
$CaCO_3$ (active)	7.22
Organic matter (Walkley Black; %)	1.76
N (total; Kijeldhal; %)	0.98
P_2O_5 ass. (Olsen; ‰)	0.019
K_2O exchang. (Dirks-Sheffer; ‰)	0.022
Chlorides (NaCl; ‰)	0.016
Na^+ soluble (‰)	0.011
Fe^{++} soluble (‰)	0.008
Mg^{++} (‰)	0.018
Electric conductivity (1:5; mS cm^{-1})	0.12
Cation exchange capacity (meq/100)	240

36 kg ha^{-1} N as ammonium diphosphate (18/46) were supplied, followed by 46 kg ha^{-1} of N in the ureic form that was spread at full tillering stage. In both years, sowing was made manually, employing 400 m^{-2} viable seeds. The 16 populations were arranged in a randomized complete block design (RCBD) with three replications; each elementary plot sized 3.0 × 3.5 m = 10.50 m^2 and was formed by 10 rows 30 cm apart. The crop was rainfed both years, as customary for cereals in many Mediterranean environments. The trend of rainfall and temperatures in the growth cycle of both years is shown in Figure 2.

In those years, as it is typical for the trial environment, the mean temperatures were generally mild (rarely below 0°C) in winter and higher than 30°C in summer (June–September); the precipitations were mostly distributed throughout the winter months, with prolonged dry periods in summer and spring. The 40 years (1978–2017) rainfall average value was 480 mm $year^{-1}$; rainfall amount was similar in the first trial year, but higher in 2015. Within the timeframe of oat growth (from February to June), 210-mm rainfall was recorded in 2014 and 55 mm in 2015.

2.3 | Observations on plants

In both years, the major morphobiological characters (descriptors) were measured in the evaluated populations as described in the UPOV guidelines Community Plant Variety Office—Office Communautaire des Varietes Vegetales (CPVO-OCVV) (rif. CPVO-TP/020/2) (Table 3). In all populations, a sample of 10 randomly selected plants, representative of general plot conditions, was taken per replication and per year, so achieving 960 individual cases for each measured variable.

In the field, each characteristic was measured when oat plants were at the optimum development stage for assessment. Twenty-one characters were considered, related to various aspects of the entire plant, stem, leaves, panicle, glumes, and grains (including primary grain; Table 3). Six out of 21 were quantitative, and all the others were qualitative traits.

2.4 | Statistical methods

Because of the mixed nature of the dataset, including both quantitative and qualitative variables, the analysis was carried out in the following steps. A first exploratory analysis was based only on continuous variables; after being standardized in order to make them comparable, as they were expressed in different units, a multivariate analysis of variance (MANOVA; Hardle & Simar, 2019) was applied to determine the effects of years, populations, and their interactions on the vector of the mean values. All the assumptions for a correct interpretation of the test results were verified: observations had been randomly and independently sampled from the populations, and the variables were linearly correlated conditionally to the years. The homogeneity of variances was assessed through Levene's test (p-values > 0.05). The null hypotheses were tested by means of the decomposition of the deviances and using the Pillai–Bartlett Trace test (Hardle & Simar, 2019); to find out which specific populations means, compared with each other, were different, Tukey post hoc test was applied (Hardle & Simar, 2019). The 6 × 6 matrix of pairwise correlation among variables was computed to analyze their mutual linear relationships, and a PCA was applied to reduce dimensionality of variables with a minimal loss of information (Jolliffe, 2002). The required number of principal components (PCs) was determined as the smallest value for which a cumulative percentage of total variation >80% was reached. Hierarchical cluster analysis for classifying the populations in a hierarchy of clusters was applied to find groups of populations according to their similarities (Everitt et al., 2011). Cluster analysis was performed through the agglomerative hierarchical clustering, and the criterion for choosing the pair of clusters, to merge at each step, was the Ward's method based on the minimum variance. A silhouette measure (s_i) quantified the cohesion of each population to its own cluster compared to the separation from the other clusters. It was computed by $s_i = (b_i - a_i)/\max(b_i, a_i)$, where a_i is the mean distance between the ith population and all others in the same cluster (the smaller the value, the better the

TABLE 3 Morphobiological qualitative and quantitative characters measured in 2014 and 2015 on oat (*Avena sativa*) populations at Sparacia (Cammarata, AG, Italy).

Plant							
1. PLHAB—Growth habit (at tillering; QL)		**2. PLHGT—Length (at seed dough development, cm, panicle included; QN)**					
1	Erect	<72.4					
3	Semi-erect	72.5–90.8					
5	Intermediate	90.9–109.2					
7	Semi-prostrate	109.3–127.6					
		>127.6					
Stem							
3. STHUN—Presence/intensity of hairiness of uppermost node (at anthesis; QL)							
0	Absent or very weak						
3	Weak						
5	Medium						
7	Strong						
9	Very strong						
Leaves							
4. LVHSH—Lowest leaves: hairiness of sheaths (at tillering; QL)		**5. LVHML—Leaf blade: hairiness of margins of leaf below flag leaf (at stem elongation; QL)**		**6.LVFLH—Flag leaf: habit (at booting; QL)**			
1	Absent or very weak	1	Absent or very weak	1	Erect		
2	Medium	3	Weak	3	Semi-erect		
3	Strong	5	Medium	5	Intermediate		
		7	Strong	7	Semi-recurved		
				9	Recurved		
Panicle							
7. PNBRO—Orientation of branches (QL)		**8. PNBRA—Attitude of branches (QL)**		**9. PNSPA—Attitude of spikelets (QL)**		**10. PNLEN—Length (cm; QN)**	
1	Unilateral	1	Erect	1	Erect	<13	
2	Sub-unilateral	3	Semi-erect	2	Drooping	13–19	
3	Divergent	5	Horizontal			19–25	
		7	Drooping			25–30	
		9	Strongly drooping				
Glumes							
11. GLGLA—Glaucosity (QL)		**12. GLLEN—Length (mm; QN)**					
1	Absent or very weak	<17.37					
3	Weak	17.37–21.74					
5	Medium	21.74–26.10					
7	Strong	26.11–30.47					
9	Very strong	>30.47					
Grain							
13. GRHSK—Husk (QL)		**14. GRLMC—Colour of lemma (QL)**					
1	Absent	1	White				
9	Present	2	Yellow				
		3	Brown				
		4	Grey				
		5	Black				

(Continues)

TABLE 3 (Continued)

Primary grain					
15. PGGLI—Intensity of glaucosity of lemma (QL)		**16. PGLLE—Length of lemma (mm; QN)**	**17. PGHBL—Hairiness of back of lemma (QL)**		**18. PGHAB—Hairiness of base (QL)**	
0	Absent	<13.6	1	Absent	1	Absent or very weak
1	Very weak	13.7–17.5	9	Present	3	Weak
3	Weak	17.6–21.3			5	Medium
5	Medium	21.4–25.2			7	Strong
		>25.2			9	Very strong

19. PGBHL—Length of basal hairs (mm; QN)	**20. PGAWN—Frequency of awns (QL)**		**21. PGRCL—Length of rachilla (mm; QN)**
0	1	Absent or low	<0.69
0.1–3.0	3	Medium	0.69–1.26
3.1–5.0	5	High	1.27–1.84
5.1–7.0	7	Strong	1.85–2.41
	9	Very strong	>2.41

Abbreviations: QL, Qualitative character; QN, quantitative character.

assignment) and b_i is the mean distance between the ith population and all the others in other clusters (the higher the value, the better the assignment). In the range between −1 and +1, a high value of s_i indicates that the object is well matched to its own cluster and indicates the appropriateness of the clustering configuration. The optimal number of clusters is estimated by maximizing the average silhouette with respect to the configuration obtained in the hierarchical clustering.

In the second step, a global analysis, based on multifactorial analysis (MFA), was performed to summarize the whole complex dataset described simultaneously by all the variables (quantitative and qualitative), considered in their structure in groups (Pagès, 2004). Adequacy of factorial analysis was assessed by means of Bartlett test (Bartlett, 1951) when the condition of multivariate normality was assumed; since this condition is not met for ordinal variables, the Kaiser–Meyer–Olkin (KMO) criterion (Cerny & Kaiser, 1977) was used to determine a correlation suitability for factor analysis. The correlation matrix was computed using Spearman's correlation on the qualitative variables measured on an ordinal scale.

Due to the high correlation characterizing the data, MFA was applied to reduce dimensionality and find groups of populations with similar profiles. The morphological characters (plant, stem, leaves, panicle, glumes, and grains including primary grain) played the role of variables of higher order, each consisting in a set of multiple variables, as specified in Table 3. A set of common factor scores were computed for populations and morphological characters in order to project them onto the principal axes and analyze communalities and discrepancies.

In this analysis, the importance of a single dimension (principal axis) is reflected by its eigenvalue, which indicates how much of the total inertia of the data is explained by this component. The quality of representation of a population (variable) on the selected axes is measured by the squared cosine: it allows to identify which populations are well projected and which variables contribute to the construction of the axes.

All statistical analyses are implemented in R language (R Core Team, 2021) and its packages FactoMineR and dendextend (Galili, 2015; Lê et al., 2008).

3 | RESULTS

3.1 | Analysis of quantitative data

Considering only the six quantitative variables (2.PLHGT, 10.PNLEN, 12.GLLEN, 16.PGLLE, 19.PGBHL, and 21.PGRCL), a MANOVA was applied as previously described to determine the effects of years, populations, and their interactions on the vector of the mean values. The results, reported in Table 4, show that all the effects are significant (p-value < $2.2e^{-16}$).

In order to assess which specific variables were significantly different between 2014 and 2015, single t-tests were also performed for the single variables; only for the panicle length (10.PNLEN) differences between years were not significant (Table 5). Hence, subsequent analyses were performed separately for both years.

In PCA, the selection of the first two components allows the projection of data from the six-dimensional space of the quantitative variables into a two-dimensional space. After this operation, 81.5% of variability was retained in the data from 2014 (the third component added up 9.9% of explained variability). In contrast, in data from 2015, the first two principal

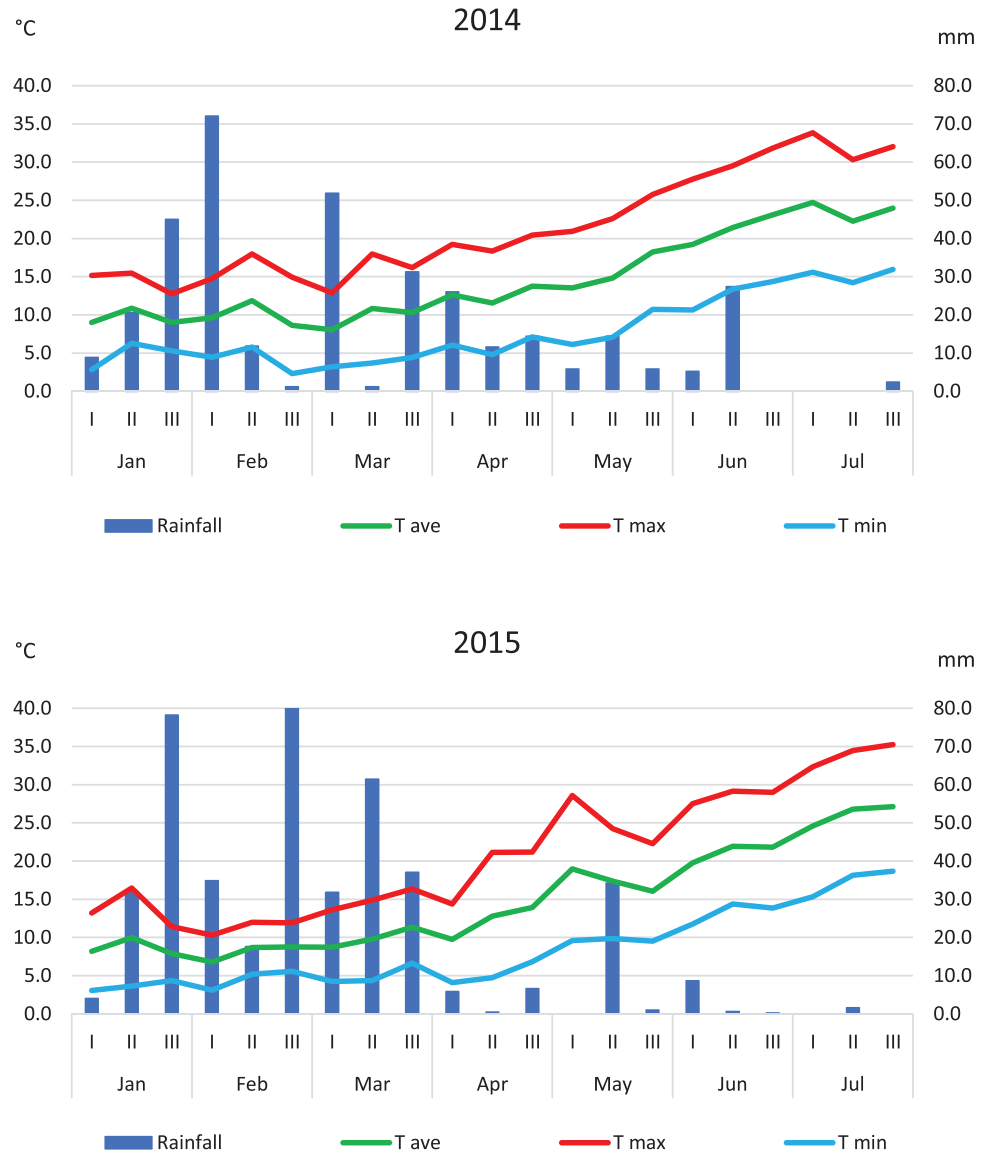

FIGURE 2 Ten days values of rainfall (mm) and temperatures (°C) recorded at "Sparacia" farm in 2014 and 2015 throughout the trials.

TABLE 4 *Avena sativa*. One-way multivariate analysis of variance (MANOVA) comparison between 2014 and 2015.

	df	Pillai-Bartlett	*F* approx	Num *df*	Den *df*	*P*(> *F*)
Year	1	0.737	430.36	6	$<2.2e^{-16}$	Year
Population	15	0.681	7.93	90	$<2.2e^{-16}$	Population
Year:Population	15	0.461	5.15	90	$<2.2e^{-16}$	Year:Population
Residuals	928					Residuals

Abbreviations: Num, numerator; Den, denominator.

components explained 71.1% of the whole variability and the third component added up 12.3% of explained variability (Table 6). Standard deviations of principal components in Table 6 are the square root of the eigenvalues. The factor loadings (Table 7) represent the weight of the variables on the first three components. Table 7 shows that each PC turns out to be a linear combination of a different subset of variables; this configuration also simplifies the interpretation of results.

In Figure 3, 16 populations are projected on the map of the first two PCs and, simultaneously, the contribution of each variable to the PCs is represented.

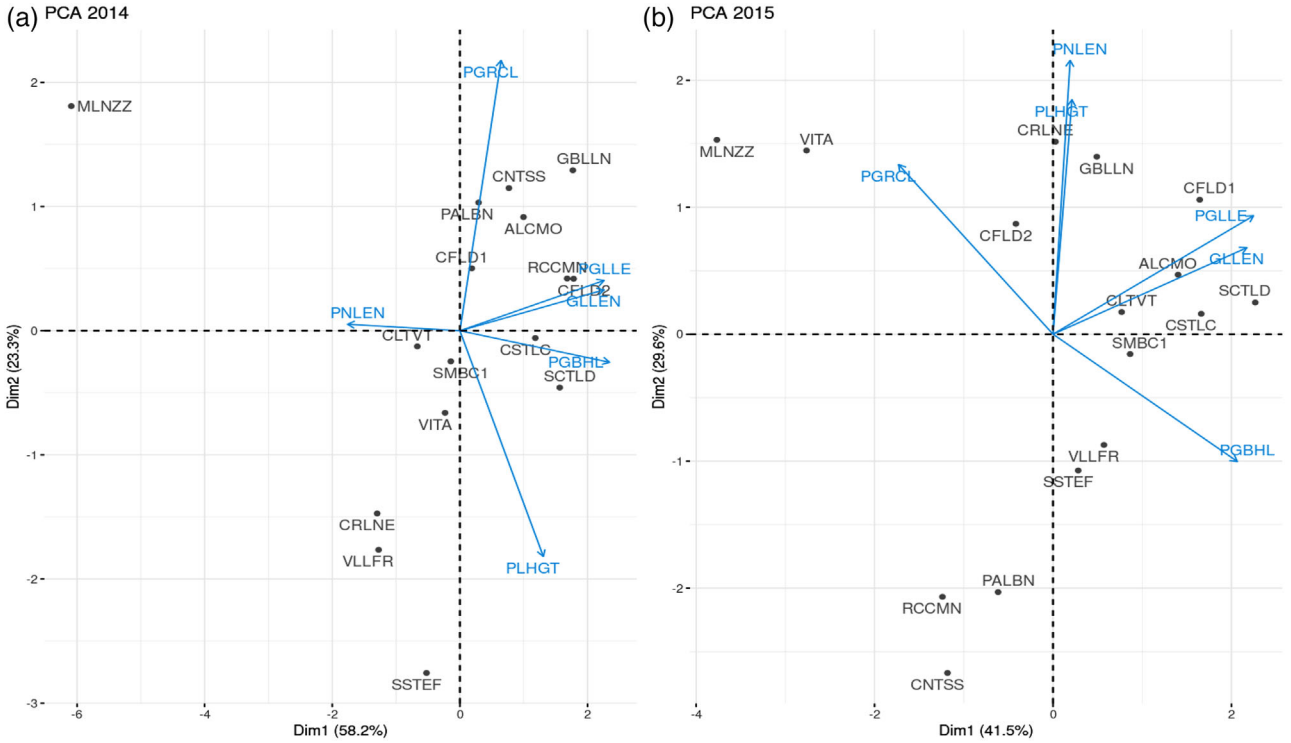

FIGURE 3 *Avena sativa* 2014–2015. Principal component analysis (PCA) based on the six quantitative variables; each population is represented by the mean of the scores, over the 30 observations in 2014 (a) and 2015 (b).

TABLE 5 *Avena sativa*. *t*-test comparing 2014 and 2015 means for each quantitative variable.

	Mean 2014	Mean 2015	*t*-Test	*p*-Value
10. PNLEN	16.85	16.55	1.47	0.101
2. PLHGT	110.59	88.99	26.46	$< 2.2e^{-16}$
12. GLLEN	24.73	23.76	4.75	$< 4.36e^{-16}$
16. PGLLE	16.4	17.19	−5.06	$< 4.36e^{-16}$
19. PGBHL	4.63	6.16	−17.12	$< 2.2e^{-16}$
21. PGRCL	1.23	0.415	28.26	$< 2.2e^{-16}$

Abbreviations: PNLEN, panicle length; PLHGT, plant length; GLLEN, glume length; PGLLE, primary grain length; PGBHL, primary grain length of basal hairs; PGRCL, primary grain length of rachilla.

TABLE 6 Characteristics of the first three principal components (PCs) representing 6 quantitative variables measured on 16 *Avena sativa* populations in 2014 and 2015.

	PC1	PC2	PC3
2014			
Standard deviation	1.870	1.183	0.768
Proportion of variance	0.582	0.233	0.098
Cumulative proportion	0.582	0.815	0.913
2015			
Standard deviation	1.579	1.332	0.859
Proportion of variance	0.410	0.295	0.123
Cumulative proportion	0.410	0.711	0.834

In the 2014 representation (Figure 3a), the first component (PC1) is mostly determined by the variables 12.GLLEN (length of glumes), 16.PGLLE (length of lemma in primary grain) and 19.PGBHL (length of basal hairs in primary grain).

The length of glumes and length of lemma were highly correlated in both years. Populations with high values on the first component (Table 7) had values of 12.GLLEN, 16.PGLLE, and 19.PGBHL higher than the global mean, and the population MLNZZ, with the lowest negative score on PC1, had the smallest values in both years. In Table 8, the reported 95% confidence intervals for CSTLC and SCTLD and for RCCMN, CFLD2, and GBLLN do not contain the global mean computed on the overall populations (Table 5).

As shown in Table 7, the variables 2.PLHGT (length of plant), with a negative sign, and 21.PGRCL (length of rachilla in primary grain), with a positive sign, strongly influenced PC2 for 2014; in Table 9, points with high values on PC2 had small PLHGT values and high PGRCL values; a negative partial correlation was also measured between the two variables ($r_{\text{PLHGT,PGRCL}} = -0.59$). The population GBLLN, having high scores on PC2 for 2014 and 2015, was characterized by high values of PGRCL in 2014 and high values of PLHGT and PNLEN in 2015.

The quality of representation of the 16 populations in the space defined by the two-dimensional factor map is computed by means of the $\cos^2$ values (Figure 4).

TABLE 7 Factor loadings on the first three principal components (PCs) of the six quantitative variables measured on 16 *Avena sativa* populations in 2014 and 2015. Scores of the most determinant variables are bolded.

	2014			2015		
	PC1	PC2	PC3	PC1	PC2	PC3
2. PLHGT	0.286	**−0.628**	0.114	0.051	**0.529**	**0.807**
10. PNLEN	−0.385	0.018	**0.896**	0.046	**0.618**	**−0.488**
12. GLLEN	**0.494**	0.114	0.307	**0.525**	0.195	−0.285
16. PGLLE	**0.494**	0.140	0.298	**0.543**	0.267	0.041
19. PGBHL	**0.513**	−0.088	0.032	**0.499**	−0.287	0.111
21. PGRCL	0.141	**0.752**	0.024	**−0.418**	0.383	−0.128

Abbreviations: PLHGT, plant height; PNLEN, panicle length; GLLEN, glume length; PGLLE, primary grain length; PGBHL, primary grain length of basal hairs; PGRCL, primary grain length of rachilla.

TABLE 8 *Avena sativa*. Confidence intervals (IC) at 95% for variables and populations most correlated with principal component 1 in 2014 and in 2015.

	2014						2015					
Location	GGLEN IC 95%		PGLLE IC 95%		PGBHL IC 95%		GGLEN IC 95%		PGLLE IC 95%		PGBHL IC 95%	
CSTLC	26.02	28.44	17.25	18.63	4.74	5.32	23.47	26.06	17.12	19.28	6.30	6.90
SCTLD	26.03	28.28	16.30	17.93	4.91	5.37	25.34	28.06	17.44	19.56	5.63	6.64
RCCMN	25.21	26.49	16.26	17.57	5.34	5.94						
CFLD2	24.47	26.71	16.54	17.69	4.93	5.55						
GBLLN	24.42	27.60	17.33	18.85	4.73	5.25						
MLNZZ	18.67	20.78	12.74	13.98	1.48	2.44	19.72	22.81	15.07	17.27	4.40	5.73

Abbreviations: GGLEN, glume length; PGLLE, primary grain length; PGBHL, primary grain length of basal hairs; CSTLC, Castel di Lucio (ME); SCTLD, San Cataldo (CL); RCCMN, Roccamena (PA); CFLD2, Cefalà Diana 2 (PA); GBLLN, Gibellina (TP); MLNZZ, Mulinazzo (AG).

TABLE 9 *Avena sativa*. Confidence intervals (IC) at 95% for variables and population most correlated with principal component 2 in 2014 and in 2015.

	2014				2015			
Location	PGRCL IC 95%		PLHGT IC 95%		PLHGT IC 95%		PNLEN IC 95%	
GBLLN	1.63	1.90	105.95	116.32	87.48	95.52	16.09	18.31

Abbreviations: PGRCL, primary grain length of rachilla; PLHGT, plant height; GBLLN, Gibellina (TP).

If points are well represented, the value of $\cos^2$ is close to one, whereas for some populations (SMBC1 in 2014 and CFLD2, SMBC1, CSTVT in 2015), the third dimension (basically the length of panicle) was required to represent them properly. Interestingly, the panicle length was the most characterizing variable in these three populations.

In the 2015 representation, populations had different positions in the space of the first two components (71.1% of explained variance), that is, different weights on the two components with respect to 2014 (Table 6; Figure 3b). Some populations (CNTSS, GBLLN, PALBN, RCCMN, SSTEF, MLNZZ, and VITA) had a better representation considering the second dimensions (Figure 4b), characterized in 2015 by a positive correlation with both variables 2.PLHGT (length of plant) and 10.PNLEN (length of panicle). Other populations (ALCMO, CLTVT, CFLD2, and SMBC1) were better represented when the third dimension was also considered; PC3 showed a positive correlation with 2.PLHGT, and a negative correlation with 10.PNLEN (Table 7). Hence, these populations were characterized by high plant length values and small panicle length values. Also, in 2015, the variable 21.PGRCL (length of rachilla in primary grain) was negatively correlated, in terms of marginal linear correlation, with all the other variables in the first PC. The partial correlations were negative as well: the partial correlation coefficient between 21.PGRCL and 12.GLLEN (length of glumes) was −0.51; between 21.PGRCL and 19.PGBH (length of basal hairs in primary grain) the correlation was −0.36, and there was no significant partial correlation between 21.PGRCL and 16.PGLLE (length of lemma of primary grain).

Some characteristics persisted over the years: factor loadings (Table 7) indicated PC1 as a linear combination of

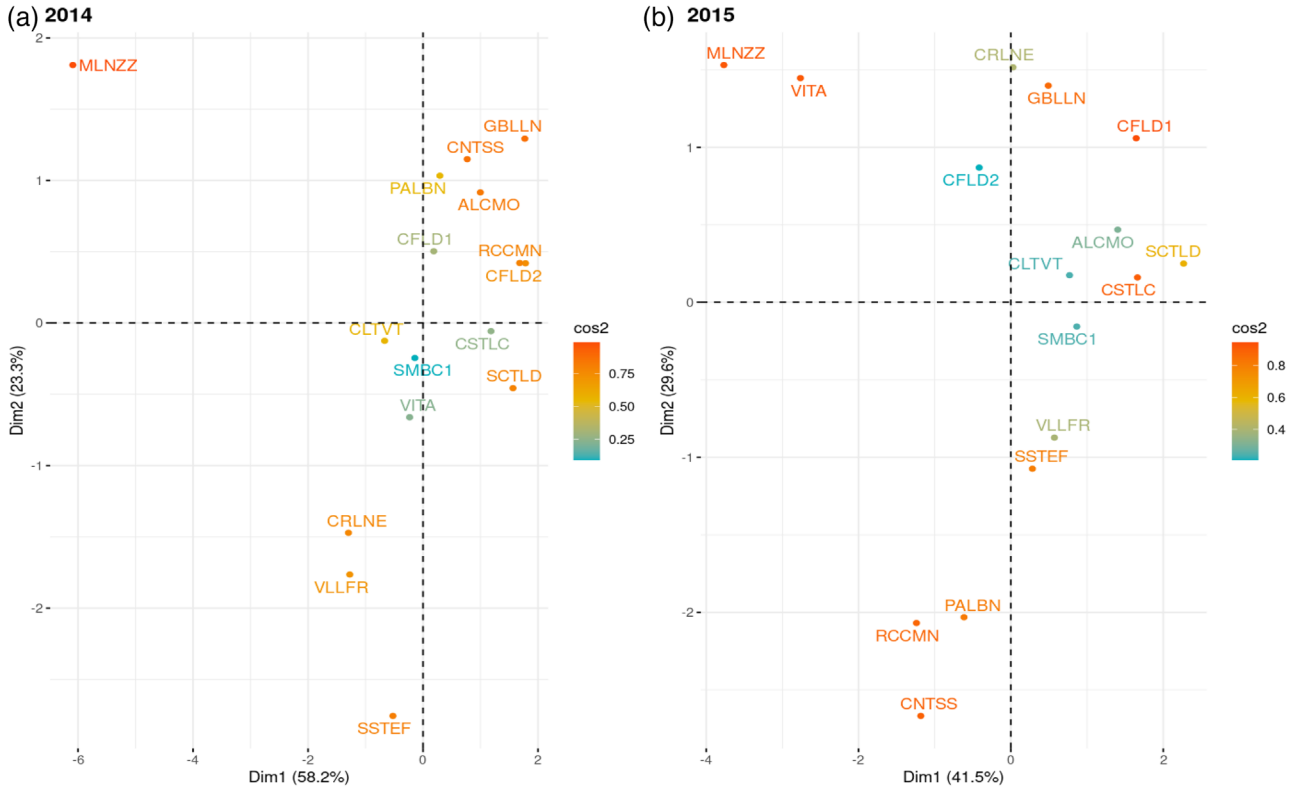

FIGURE 4 Cos^2 values of the 16 *Avena sativa* studied populations, represented on the factor map in 2014 (a) and 2015 (b).

12.GLLEN (length of glumes), 16.PGLLE (length of lemma of primary grain), and 19.PGBHL (length of basal hairs in primary grain). Among the populations, the profile of MLNZZ is very distant from the others.

The optimal groups of populations in 2014 and 2015, resulting from cluster analysis (Figure 5), are reported in Table 10 (cluster membership) and Table 11 (silhouette measures as degree of appropriateness of the clustering configuration, that is, null in clusters with one element).

Evidence for the presence of clusters of observations can often be found in one or two dimensions, by first projecting the data into the lower dimensional space of the first two principal components (Factor Map).

For both years, the principal components properly synthesize the clusters effects in R^d and this synthesis can be easily visualized in the subspace spanned by the first two principal components ($d = 2$), where the obtained clusters are well separated (Figure 6).

The differences in distribution between characters, observed in 2014 and 2015, also imply different results in CA.

Implementing the agglomerative hierarchical clustering, with Ward's method as the criterion of aggregation, and selecting the optimal number of clusters based on the maximum average silhouette width, the optimal configuration obtained for 2015 consisted of four groups (Tables 10 and 11; Figure 5b) that were well separated on the subspace of the first two PC (Figure 6b).

As further confirmation of the cohesiveness of the obtained clusters and the significant results in Table 5, Tukey post hoc test (results not reported) was performed to compare differences between all the possible pairs of population means, for separated years; for 2014, the test produced significant results only among populations resulting as members of different clusters; for 2015, only in three comparisons, means of 21.PGRCL were found significantly different for populations belonging to the same cluster, that is, CRLNE-SCTLD, CRLNE-CSTLC, and CRLNE-CFLD1.

3.2 | Multifactorial analysis on mixed data

In the second step of the analysis, the categorical variables were included, implementing a MFA, where the six morphological characters were considered in terms of multiple variables (measurements made on entire plant, stem, leaves, panicle, glumes, and grains—including primary grain), as specified in Table 3. This allowed evaluating the importance of the variables of higher order by comparing the results obtained using mixed data with those from the first step using quantitative data.

Suitability for the methodology applied was assessed by KMO criterion for the overall data, KMO = 0.67.

The results of the MFA, in terms of morphological characters (Table 12) show that the first two dimensions explained

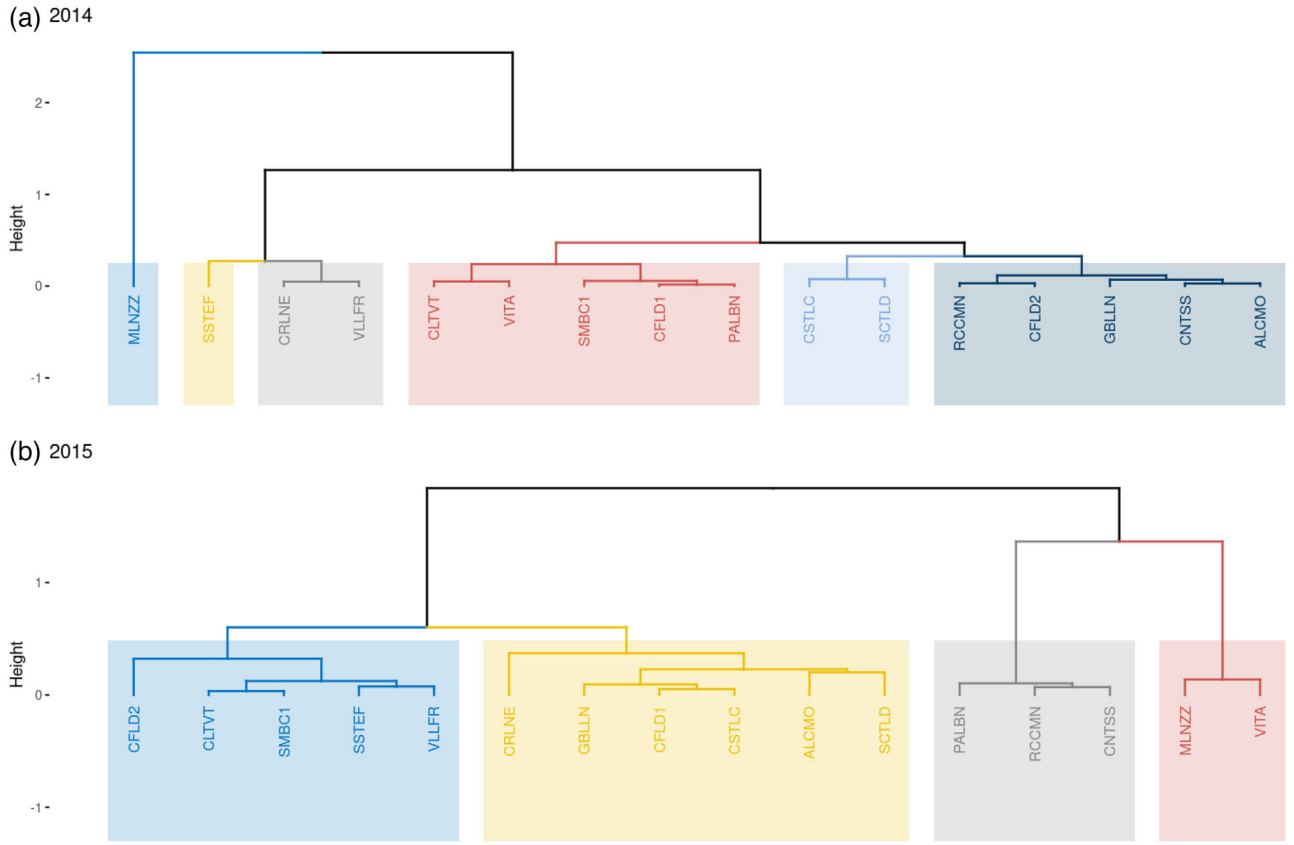

FIGURE 5 Principal component analysis (PCA) cluster membership of the 16 *Avena sativa* studied populations in 2014 (a) and 2015 (b).

FIGURE 6 Clusters projected on the space of the first two principal components (PCs) for the 16 *Avena sativa* studied populations in 2014 (a) and 2015 (b).

TABLE 10 *Avena sativa* 2014–2015. Clusters of populations based on the six considered quantitative variables.

	ALCMO	CFLD2	CNTSS	GBLLN	RCCMN	CFLD1	CLTVT	PALBN	SMBC1	VITA	CRLNE	VLLFR	CSTLC	SCTLD	MLNZZ	SSTEF
2014	1	1	1	1	1	2	2	2	2	2	3	3	4	4	5	6
2015	1	2	4	1	4	1	2	4	2	3	1	2	1	1	3	2

Abbreviations: ALCMO, Alcamo (TP); CFLD2, Cefalà Diana 2 (PA); CNTSS, Contessa Entillina (PA); GBLLN, Gibellina (TP); RCCMN, Roccamena (PA); CFLD1, Cefalà Diana 1 (PA); CLTVT, Caltavuturo (PA); PALBN, Piana Degli Albanesi (PA); SMBC1, Sambuca 1 (AG); VITA, Vita (TP); CRLNE, Corleone (PA); VLLFR, Villafrati (PA); CSTLC, Castel di Lucio (ME); SCTLD, San Cataldo (CL); MLNZZ, Mulinazzo (AG); SSTEF, Santo Stefano Quisquina (AG).

a small amount of variability (10% for 2014 and 8.11% for 2015), and a higher number of dimensions was required for achieving a suitable representation.

Total inertia measures the overall variation or differences in the populations (or categories) profiles and is the sum of the eigenvalues that are interpretable as the decomposition of the inertia along the principal axes. Figure 7 shows the contributions of the characters to the construction of the two-dimensional representation on the factor map: measurements on glumes and grain, with the highest value on the first two dimension in both years, provided a better characterization of the populations; measurements on the entire plant were relevant only in 2014, and measurements on panicle in 2015.

The low percentage of variability explained in the 2 years made necessary to consider the contribution of all the variables in other dimensions; moreover, the eigenvalues of the first four dimensions (17% for 2014 and 15.5% for 2015) were decomposed across populations or across variables, determining the respective contributions to inertia; these contributions, reflecting the proportion of the variance of a dimension that can be attributed to the variables, give diagnostics to determine the drivers of the dimensions. The larger the contribution to a dimension, the more important this variable for the dimension. The same goes for populations, and proximity among them implies similar profiles.

Histograms in Figure 8 show the contributions of the descriptors related to glumes, grain, and entire plant to the definition of the first four dimensions (17% of the whole variability) for 2014. It is possible to observe that (a) the variability linked to the glumes, grain, and entire plant contributed to the definition of the first dimension (5.51%); (b) the variability linked to the glumes contributed to the definition of the second dimension (4.34%); (c) stem and panicle contributed to the third dimension (3.99%); and (d) glumes and leaves contributed to the fourth dimension (3.62%).

In 2015 (Figure 9), glumes contributed to the first four dimensions (4.49%, 3.63%, 3.56%, and 3.36%); only grain gave an additional contribution to the first dimension and stem to the third one. As concerns the relevance of variables and categories, Figure 10 shows the most relevant ones, and their contributions to the first four dimensions.

In 2014 (Figure 10a), the highest contributions were from:

(a) glume lengths, mostly ranging from lowest values (<17.37 mm) to 21.74 mm;
(b) highest and lowest values of 6.LVFLH (LVFLH: 9; LVFLH: 1), meaning that the predominant flag leaf habit at booting was "recurved" or "erect";
(c) plant height, from the lowest values (<72.4 cm) to 90.8 cm and >127.6 cm.

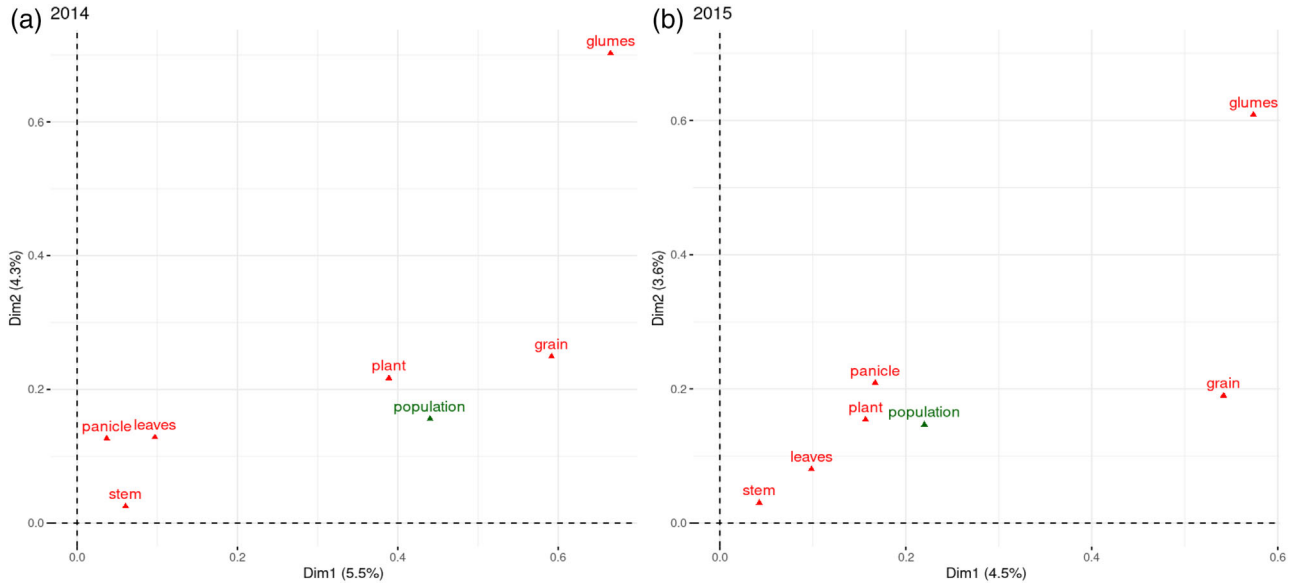

FIGURE 7 Multifactorial analysis (MFA) on the 16 *Avena sativa* studied populations. Representation of the morphological characters on the first two dimensions.

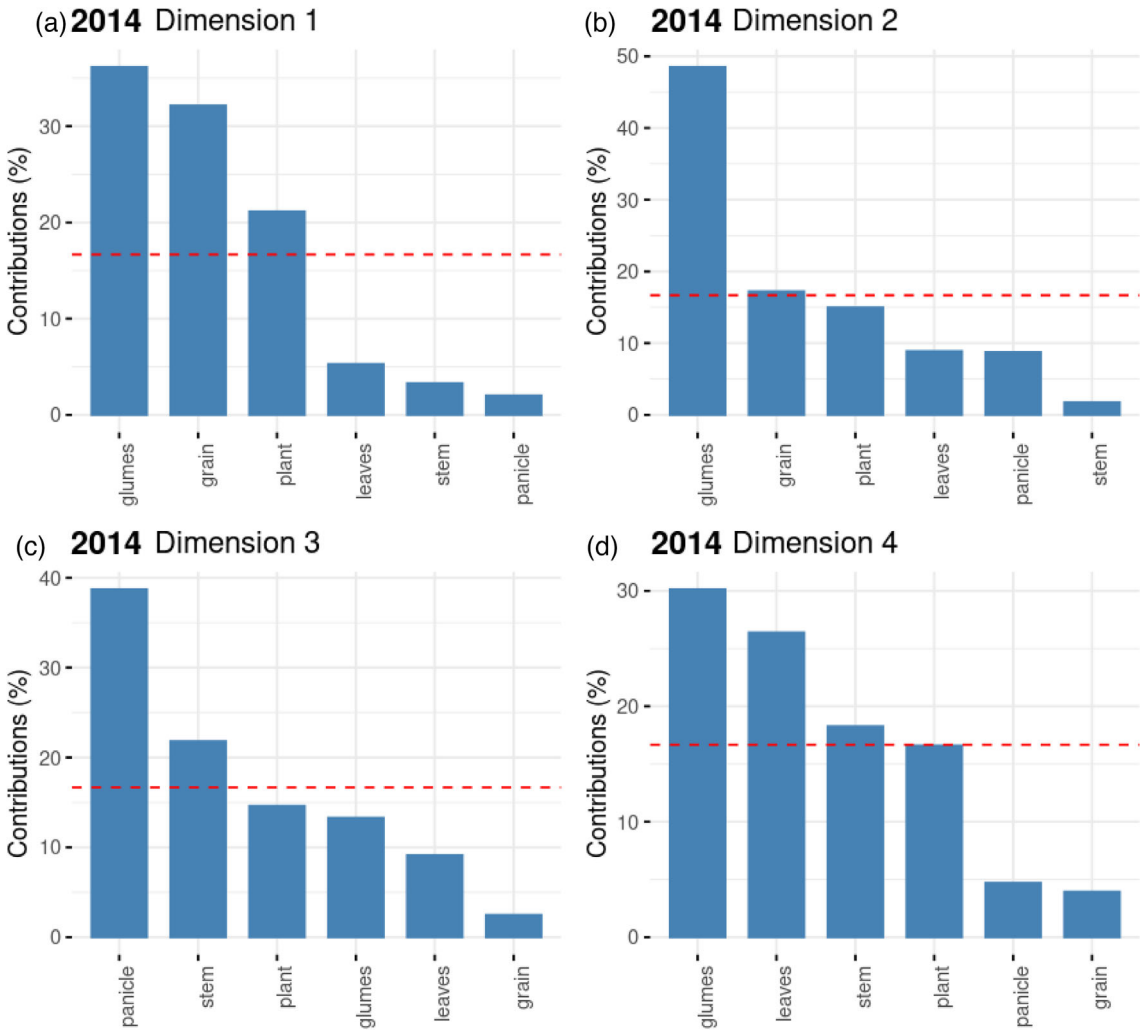

FIGURE 8 Multifactorial analysis (MFA) on the 16 *Avena sativa* studied populations contribution of the characters to the first four dimensions in 2014.

TABLE 11 *Avena sativa* 2014–2015. Silhouette width of the clusters.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
2014	0.51	0.62	0.67	0.52	0	0
2015	0.37	0.53	0.68	0.75	–	–

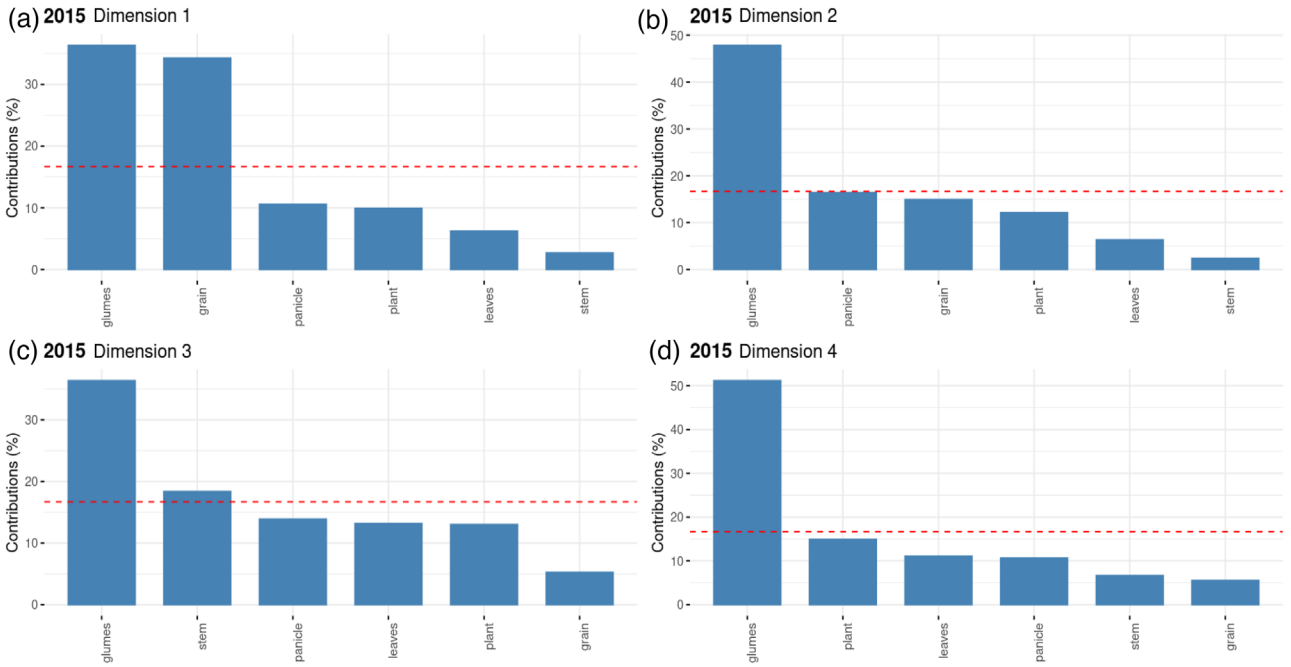

FIGURE 9 Multifactorial analysis (MFA) on the 16 *Avena sativa* studied populations. Contribution of the characters to the first four dimensions in 2015.

Lower values for descriptors of panicle (PNBRA:1, i.e., panicle branches erect, and PNBRO:1; i.e., unilateral orientation of panicle branches) and stem, no hairiness in plant's uppermost node (STHUN:1), give a weak contribution.

Additional details are given as concerns the populations: for example, the most contributive categories of glume length (<17.37 mm; 17.37–21.74 mm) and plant height (<72.5 cm; 72.5–90.8 cm) were observed in the populations CFLD2, CFLD1, MLNZZ, VITA, CLTVT, and SMBC1, indicating similar profiles of these populations, even if this interpretation is constrained by the poor explanation of inertia of the first four dimensions (17.0%), and a more in-depth analysis is required.

In 2015 (Figure 10b), the MFA analysis indicated that the contribution of descriptors related to the glumes on the principal factors was higher. The higher contributions are from all the categories of glume length, whereas values for descriptors of length and ramification attitude of panicle (PNLEN: 25–30, PNBRA:2), height of plant (PLHGT from 99.9 cm to 127.6 cm), characters of stem (absence of hairiness in the uppermost node STHUN:1), and leaves (LVFLH: 1 e LVFLH: 5) give weak contribution. The reduction of inertia explained by the projection of the points in the subspace of the first two principal dimensions (8.11%, Table 5) can be interpreted as the presence of more noise in the data, making the synthesis more difficult.

After the insertion of qualitative variables, the clustering algorithm was performed again for both years, based on the coordinates of the populations on the space of the first 20 dimensions (Figure 11). The rate of inertia explained by the first 20 dimensions was 63% for 2014 and 59% for 2015.

For 2014 data, a new configuration was obtained (Figure 11a), and the differences with respect to the cluster configuration exclusively based on the quantitative variables (Figure 5a) could be analyzed. When the categorical variables were included in the analysis, the link between some couples remained stable: CSTLC–SCTLD, CRLNE–VLLFR, SMBC1–CLTVT, CFLD2–ALCMO, and RCCMN–CNTSS. Contrastingly, differences emerged for SSTEF, grouped together with CRLNE–VLLFR and SMBC1–CLTVT; GBLLN and CFLD1, which formed a new cluster; VITA was clustered with the couple CFLD2–ALCMO; PALBN was clustered with the couple RCCMN–CNTSS.

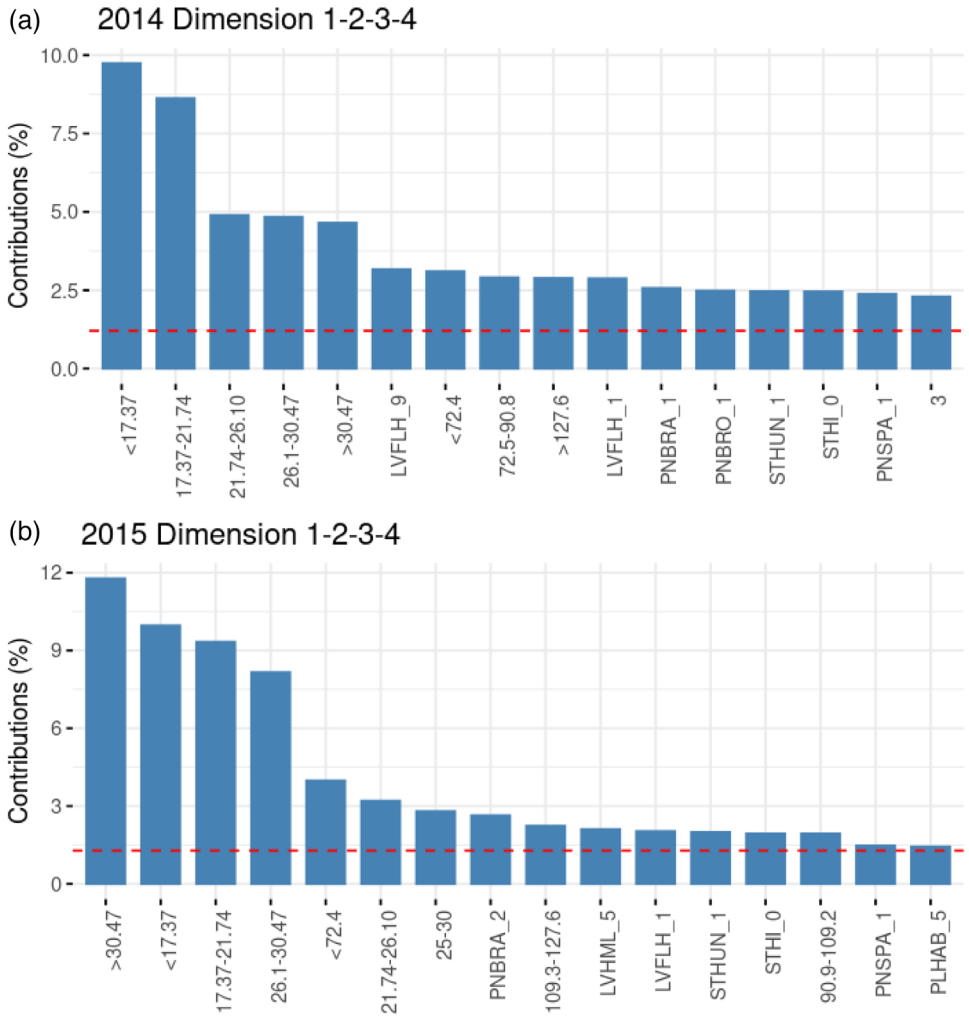

FIGURE 10 Multifactorial analysis (MFA) on the 16 *Avena sativa* studied populations. Contribution of the descriptors to the first four dimensions in 2014 (a) and 2015 (b).

Comparing the results of 2015, for quantitative data (Figure 5b) and for quantitative and categorical data (Figure 11b), differences were observed for GBLLN and CFLD1, forming a new cluster as in 2014, after having left CRLNE–CSTLC–SCTLD–ALCMO that merged with CFLD2–VLLFR; the cluster PALBN–RCCMN–CNTSS merged with CLTVT–SMBC1 and SSTEF. The cluster VITA–MLNZZ was stable.

4 | DISCUSSION AND CONCLUSIONS

The individuation of effective and straightforward methods for the description and classification of oat landraces is of the utmost importance for a proper evaluation of local germplasm. Yet, most of categorical descriptors reported in the official descriptors lists are ruled out from research on intraspecific variability, and all related analyses, including MA, are performed only on continuous data. With the purpose of offering an additional point of view, our analysis was performed considering both continuous and categorical descriptive data.

As concerns our first goal, that is, to analyze the degree of difference/similarity among the observed populations, the statistical analysis highlighted a strong variability between years for most of the examined variables. In many cases, such variability was higher than the variability between populations. Hence, the UPOV requirements for "distinctness, uniformity and stability" (UPOV, 2018) were not satisfied, and a complete, and stable over years, characterization of genotypes was not always possible. As a matter of fact, the occurrence of high levels of seasonal variability is very common in Mediterranean environments, where erratic climatic conditions can cause significant differences in plants' features (Arnon, 1992). From an evolutionary point of view, the high variability in the expression of the morphobiological characters of oat is reasonably one of the leading traits of their acknowledged rusticity toward cropping environment and energetical inputs. A high variability in cultivated

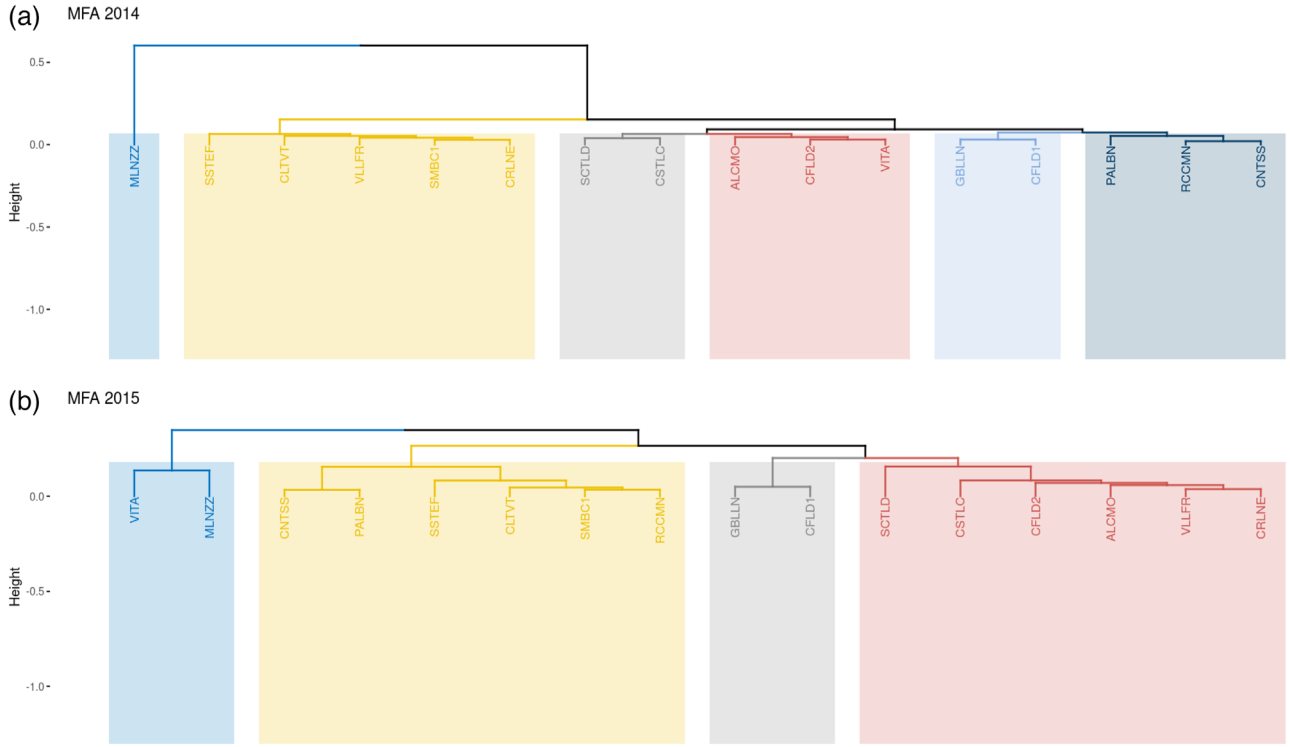

FIGURE 11 Multifactorial analysis (MFA) clusters configuration of the 16 *Avena sativa* studied populations in 2014 (a) and 2015 (b).

autochthonous populations is probably due to the high number of utilizations, alone or combined, which addressed farmers' efforts to select populations suitable for specific uses (Alonso-Blanco et al., 2005; Kalisz & Kramer, 2008) (Table 1).

Our second goal was to assess the relative importance of the morphobiological descriptors for full characterization of oat populations. Although additional plant traits (e.g., grain or biomass yield) should have been helpful for a complete characterization of oat populations, the analyses carried out on morphological data allowed some interesting considerations.

A different configuration of clusters and a different importance of variables were found when considering two different approaches: PCA based only on continuous variables and MFA based on a larger set of mixed variables, both categorical and continuous. In the second case, the chosen approach reduces information on data, as all the variables are analyzed through the decomposition of the total inertia in terms of contribution of populations and variables; this implies a larger base of data, the chance of considering a sort of "aggregates of variables", that is, all traits measured on plant, grain, stem, leaves, panicle, and glumes, having a natural interpretation for the dataset, but a reduction of the quality of the representations. In this comparison interesting differences can be appreciated, while partial results are confirmed. In particular, a higher presence of noise in 2015 affects both analyses, PCA and MFA. The four configurations obtained also reveal that the behavior of some populations is confirmed across the years (different conditions) or across the analyses (different methodologies); for example, the couples CSTLC–SCTLD and RCCMN–CNTSS are always found in the same cluster. Interestingly, the populations of the first couple were collected from livestock farms where the double utilization (grazing + hay) was prevailing, and the remaining two (RCCMN–CNTSS) came from farms mainly addressed to the production of grain (Table 1). For optimal grazing, oat must have quick early growth and a good aptitude to regrow after grazing; although all accessions in experimental fields were cultivated with an identical technique, and no efforts were made to stress or evaluate these aspects of plant performance, it seems likely that farmers acted as driving force, achieving the goal of selecting similar populations for similar uses.

Likewise, the group RCCMN–CNTSS–PALBN and the couple GBLLN–CFLD1 were observed in three of the considered configurations (PCA-2015, MFA-2014, and MFA-2015), and the group SMBC1–CLTVT–SSTEF was in the configurations of PCA-2015, MFA-2014, and MFA-2015, although in PCA-2014, the population SSTEF behaved very differently from all the other populations.

Other stable couples, detectable in three configurations out of four, were VLLFR–CRLNE and ALCMO–CFLD2, both in PCA-2014, MFA-2014, and MFA-2015.

It was not always easy to find the reason underlying these groupings: in most cases, the populations allocated in the same group, hence showing the highest degree of similarity, were gathered from rather close farms, meaning that,

TABLE 12 *Avena sativa*. Eigenvalues and percentage of explained inertia of the multifactorial analysis (MFA) on data from 2014 and 2015. Dim, dimension.

	Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	Dim 9	Dim 10	Dim 11	Dim 12	Dim 13	Dim 14	Dim 15
2014															
Inertia ratio(%)	0.325	0.269	0.234	0.215	0.207	0.208	0.198	0.200	0.192	0.191	0.182	0.178	0.181	0.176	0.174
Eigenvalue(%)	5.50	4.34	3.99	3.62	3.39	3.38	3.32	3.22	3.11	3.00	2.88	2.82	2.74	2.69	2.62
Eigenvalue cumulative(%)	5.50	9.84	13.83	17.45	20.84	24.22	27.54	30.76	33.87	36.87	39.75	42.57	45.31	48.00	50.62
2015															
Inertia ratio(%)	0.278	0.238	0.227	0.217	0.208	0.216	0.210	0.204	0.192	0.195	0.180	0.189	0.181	0.178	0.176
Eigenvalue(%)	4.49	3.62	3.56	3.36	3.27	3.25	3.13	3.026	2.94	2.84	2.76	2.72	2.63	2.57	2.45
Eigenvalue cumulative(%)	4.49	8.11	11.68	15.05	18.33	21.59	24.73	27.75	30.69	33.53	36.30	39.03	41.66	44.24	46.70

most probably, they belong to the same genotype. That was the case for the couple RCCMN–CNTSS, and to a certain extent also PALBN. It is arguable that additional observations (e.g., the grain yield, or time to anthesis, or tillering aptitude) could enlighten some mechanisms that the simple morphological data cannot disclose. In this sense, morphological data alone do not give exhaustive information, and there is room for further analyses keeping into account other characters more correlated with the possibility of use of the given genotypes.

From the methodological standpoint, it is possible to conclude that the data collected pose the challenge of dealing with multidimensional mixed data. The applied methodologies offer an in-depth analysis of the morphological aspects and their descriptors. A possible development is to consider an alternative approach that attempts to save the advantages of MFA, avoiding the loss of information of quantitative data.

AUTHOR CONTRIBUTIONS

Alessandra Carrubba: Conceptualization; data curation; formal analysis; investigation; methodology; visualization; writing—original draft; writing—review and editing. **Francesca Di Salvo**: Conceptualization; formal analysis; investigation; methodology; visualization; writing—original draft; writing—review and editing. **Roberto Marceddu**: Data curation; investigation; methodology; visualization; writing—original draft; writing—review and editing. **Mauro Sarno**: Conceptualization; data curation; funding acquisition; project administration; resources; supervision; validation; visualization; writing—review and editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

Adewale, D. B., Adeigbe, O. O., Adenuga, O. O., Adepoju, A. F., Muyiwa, A. A., & Aikpokpodion, P. O. (2013). Descriptive and discriminatory significance of pod phenotypic traits for diversity analysis of cocoa genotypes. *Journal of Plant Breeding and Genetics*, *1*(3), 131–137.

Alonso-Blanco, C., Mendez-Vigo, B., & Koornneef, M. (2005). From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. *International Journal of Developmental Biology*, *49*, 717–732. https://doi.org/10.1387/ijdb.051994ca

Arnon, I. (1992). *Agriculture in drylands: Principles and practices*. Elsevier.

Bartlett, M. S. (1951). A further note on tests of significance in factor analysis. *British Journal of Psychology*, *4*, 1–2.

Bono, G., Di Pisa, A., Giambalvo, D., & Di Miceli, G. (2007). Risposta produttiva di popolazioni e varietà di avena al variare della modalità di gestione. In L. S. Cosentino & R. Tuttobene (Eds.), *Proceedings of XXXVII Convegno della Società Italiana di Agronomia* (pp. 101–102). Università di Catania. (In Italian).

Bono, G., Giambalvo, D., & Crescente, F. (2008). Caratterizzazione bioagronomica di popolazioni siciliane di avena (*Avena sativa* L.). In L. De Bellis, S. Marchiori & A. Miceli (Eds.), *Proceedings of VIII Convegno Nazionale sulla Biodiversità* (pp. 74–76). Università del Salento. (In Italian).

Bono, G., Gristina, L., Mazza, F., & Poma, I. (2009). Valutazione bioagronomica e qualitativa di germoplasma autoctono di *Avena sativa* L. in Sicilia. In G. Sarli, A. Alvino & C. Cervelli (Eds.), *Proceedings of IV Convegno Nazionale Piante Mediterranee* (pp. 154–161). (In Italian).

Butt, M. S., Tahir-Nadeem, M., Khan, M. K. I., Shabir, R., & Butt, M. S. (2008). Oat: Unique among the cereals. *European Journal of Nutrition*, *47*(2), 68–79. https://doi.org/10.1007/s00394-008-0698-7

Caramante, M., Rouphael, Y., & Corrado, G. (2021). The genetic diversity and structure of tomato landraces from the Campania Region (Southern Italy) uncovers a distinct population identity. *Agronomy*, *11*, 564. https://doi.org/10.3390/agronomy11030564

Carrubba, A., Abbate, L., Sarno, M., Sunseri, F., Mauceri, A., Lupini, A., & Mercati, F. (2020). Characterization of Sicilian rosemary (*Rosmarinus officinalis* L.) germplasm through a multidisciplinary approach. *Planta*, *251*, Article 37. https://doi.org/10.1007/s00425-019-03327-8

Castañeda-Garzón, S. L., Ligarreto, M. G. A., & Miranda-Lasprilla, D. (2016). Morphological evaluation of an in situ collection of species from the *Annonaceae* family in Colombia. *Agronomía Colombiana*, *34*(2), 131–140. https://doi.org/10.15446/agron.colomb.v34n2.54527

Cerny, B. A., & Kaiser, H. F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research*, *12*(1), 43–47. https://doi.org/10.1207/s15327906mbr1201_3

de Oliveira, E. J., de Oliveira Filho, O. S., & da Silva, S. V. (2014). Selection of the most informative morphoagronomic descriptors for cassava germplasm. *Pesquisa Agropecuária Brasileira*, *11*(49), 891–900. https://doi.org/10.1590/S0100-204X2014001100008

Dumlupinar, Z., Dokuyucu, T., Maral, H., Kara, R., & Akkaya, A. (2012). Evaluation of Turkish oat landraces based on morphological and phenological traits. *Žemdirbystė = Agriculture*, *99*(2), 149–158.

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

FAOSTAT (2022). Crops and livestock products. http://www.fao.org/faostat/en/#data/QC

Fiore, M. C., Mercati, F., Spina, A., Blangiforti, S., Venora, G., Dell'Acqua, M., Lupini, A., Preiti, G., Monti, M., Pè, M. E., & Sunseri, F. (2019). High-throughput genotype, morphology, and quality traits evaluation for the assessment of genetic diversity of wheat landraces from Sicily. *Plants*, *8*, 116. https://doi.org/10.3390/plants8050116

Francia, E., Pecchioni, N., Destri Nicosia, O. L., Paoletta, G., Taibi, L., Franco, V., Odoardi, M., Stanca, A. M., & Delogu, G. (2006). Dual-purpose barley and oat in a Mediterranean environment. *Field Crops Research*, *99*(2–3), 158–166. https://doi.org/10.1016/j.fcr.2006.04.006

Galili, T. (2015). dendextend: An R package for visualizing, adjusting, and comparing trees of hierarchical clustering. *Bioinformatics*, *31*(22), 3718–3720. https://doi.org/10.1093/bioinformatics/btv428

Hardle, W. K., & Simar, L. (2019). *Applied multivariate statistical analysis*. Springer. https://doi.org/10.1007/978-3-030-26006-4

Ihsan, M., Nazir, N., Ghafoor, A., Khalil, A., Zahoor, M., Nisar, M., Khames, A., Ullah, R., & Shah, A. (2021). Genetic diversity in local and exotic *Avena sativa* L. (oat) germplasm using multivariate analysis. *Agronomy*, *11*, 1713. https://doi.org/10.3390/agronomy11091713

Jolliffe, I. T. (2002). *Principal component analysis. Springer series in statistics*. Springer.

Jones, H., Civáň, P., Cockram, J., Leigh, F. J., Smith, L. M., Jones, M. K., Charles, M. P., Molina-Cano, J.-L., Powell, W., Jones, G., & Brown, T. A. (2011). Evolutionary history of barley cultivation in Europe revealed by genetic analysis of extant landraces. *BMC Evolutionary Biology*, *11*, 320. http://www.biomedcentral.com/1471-2148/11/320

Junior e Silva, W. C., Carvalho, S. I. C., & Duarte, J. B. (2013). Identification of minimum descriptors for characterization of *Capsicum* spp. germplasm. *Horticultura Brasileira*, *31*, 190–202.

Kalisz, S., & Kramer, E. M. (2008). Variation and constraint in plant evolution and development. *Heredity*, *100*, 171–177. https://doi.org/10.1038/sj.hdy.6800939

Kumar, R., Varghese, S., Jayaswal, D., Jayaswall, K., Yadav, K., Mishra, G., Vyas, R. P., Singh, H. C., Prakash, H. G., Singh, A. N., & Kumar, S. (2023). Agro-morphological and genetic variability analysis in oat germplasms with special emphasis on food and feed. *PLoS One*, *18*, e0280450. https://doi.org/10.1371/journal.pone.0280450

Kumar, T., Bharadwa, C., Rizvi, A. H., Sarker, A., Tripathi, S., Alam, A., & Chauhan, S. K. (2015). Chickpea landraces: A valuable and divergent source for drought tolerance. *International Journal of Tropical Agriculture*, *33*(3), 1–6.

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, *25*(1), 1–18. https://doi.org/10.18637/jss.v025.i01

Le Dien, S., & Pagés, J. (2003). Analyse factorielle multiple hiérarchique. *Revue de Statistique Appliquée*, *51*(2), 83–93. http://www.numdam.org/item/RSA_2003__51_2_47_0/

Ligarreto, G. (2003). Caso 2. Análisis de la variabilidad genética en fríjol. In T. L. Franco & R. Hidalgo (Eds.), *Análisis estadístico de datos de caracterización morfológica de recursos fitogenéticos* (pp. 40–49). Boletín técnico no. 8. Instituto Internacional de Recursos Fitogenéticos (IPGRI). https://www.bioversityinternational.org/fileadmin/_migrated/uploads/tx_news/Análisis_estadístico_de_datos_de_caracterización_morfológica_de_recursos_fitogenéticos_894.pdf

Pagès, J. (2004). Multiple factor analysis: Main features and application to sensory data. *Revista Colombiana de Estadística*, *27*(1), 1–26.

Piergiovanni, A. R., & Lioi, L. (2010). Italian common bean landraces: History, genetic diversity and seed quality. *Diversity*, *2*, 837–862. https://doi.org/10.3390/d2060837

R Core Team (2021). R: A language and environment for statistical computing. R foundation for statistical computing. https://www.R-project.org/

Rojas, W. (2003). Caso 1. Análisis de la variabilidad genética en quinua. In T. L. Franco & R. Hidalgo (Eds.), *Análisis Estadístico de Datos de Caracterización Morfológica de Recursos Fitogenéticos* (pp. 27–37). Boletín técnico no. 8. Instituto Internacional de Recursos Fitogenéticos (IPGRI). https://www.bioversityinternational.org/fileadmin/_migrated/uploads/tx_news/Análisis_estadístico_de_datos_de_caracterización_morfológica_de_recursos_fitogenéticos_894.pdf

Roselló, M., Villegas, D., Álvaro, F., Soriano, J. M., Lopes, M. S., Nazco, R., & Royo, C. (2019). Unravelling the relationship between adaptation pattern and yield formation strategies in Mediterranean durum wheat landraces. *European Journal of Agronomy*, *107*, 43–52. https://doi.org/10.1016/j.eja.2019.04.003

Ruisi, P., Ingraffia, R., Urso, V., Giambalvo, D., Alfonzo, A., Corona, O., Settanni, L., & Frenda, A. S. (2021). Influence of grain quality, semolinas and baker's yeast on bread made from old landraces and modern genotypes of Sicilian durum wheat. *Food Research International*, *140*, 110029. https://doi.org/10.1016/j.foodres.2020.110029

Sánchez-Martín, J., Rubiales, D., Flores, F., Emeran, A. A., Shtaya, M. J. Y., Sillero, J. C., Allagui, M. B., & Prats, E. (2014). Adaptation of oat (*Avena sativa*) cultivars to autumn sowings in Mediterranean environments. *Field Crops Research*, *156*, 111–122. https://doi.org/10.1016/j.fcr.2013.10.018

Silva, R. S., Moura, E. F., Farias-Neto, J. T., Ledo, C. A. S., & Sampaio, J. E. (2017). Selection of morphoagronomic descriptors for the characterization of accessions of cassava of the Eastern Brazilian Amazon. *Genetics and Molecular Research*, *16*(2), gmr16029595. https://doi.org/10.4238/gmr16029595

Singh, R., De, S., & Belkheir, A. (2013). *Avena sativa* (oat), A potential nutraceutical and therapeutic agent: An overview. *Critical Reviews in Food Science and Nutrition*, *53*(2), 126–144. https://doi.org/10.1080/10408398.2010.526725

Šmídová, Z., & Rysová, J. (2022). Gluten-free bread and bakery products technology. *Foods*, *11*, 480. https://doi.org/10.3390/foods11030480

Soil Survey Staff (2014). *Keys to soil taxonomy* (12th ed.). United States Department of Agriculture - Natural Resources Conservation Service.

UPOV (2018). *Oats. Guidelines for the conduct of tests for distinctness, uniformity and stability*. https://www.upov.int/edocs/mdocs/upov/en/tc_54/tg_20_11_proj_5.pdf

Villa, T. C. C., Maxted, N., Scholten, M., & Ford-Lloyd, B. (2005). Defining and identifying crop landraces. *Plant Genetic Resources*, *3*(3), 373–384. https://doi.org/10.1079/PGR200591

Wagh, V. R., Sonone, A. H., & Damame, S. V. (2019). Assessment of genetic diversity in forage oat (*Avena sativa* L.). *Forage Research*, *45*(3), 203–205.

Weckwerth, W., Ghatak, A., Bellaire, A., Chaturvedi, P., & Varshney, R. K. (2020). PANOMICS meets germplasm. *Plant Biotechnology Journal*, *18*, 1507–1525. https://doi.org/10.1111/pbi.13372

Zhang, K., Dong, R., Hu, X., Ren, C., & Li, Y. (2021). Oat-based foods: Chemical constituents, glycemic index, and the effect of processing. *Foods*, *10*(6), 1304. https://doi.org/10.3390/foods10061304

Zwer, P. K. (2016). Oats: Overview. In C. Wrigley, H. Corke, K. Seetharaman, & J. Faubion (Eds.), *Encyclopedia of food grains* (Vol. 1, 2nd ed., pp. 173–183). Elsevier. https://doi.org/10.1016/B978-0-12-394437-5.00013-9

How to cite this article: Carrubba, A., Di Salvo, F., Marceddu, R., & Sarno, M. (2023). Variability of morphological descriptors in Sicilian oat (*Avena sativa* L.) populations. *Agronomy Journal*, *115,* 2469–2488. https://doi.org/10.1002/agj2.21423