Genetic and Morpho-Agronomic Characterization of Sicilian Tetraploid Wheat Germplasm

Angelo Sicilia
University of Catania
Umberto Anastasi
University of Catania
Michele Bizzini
Stazione Consorziale Sperimentale di Granicoltura per la Sicilia
Stefania Montemagno
Stazione Consorziale Sperimentale di Granicoltura per la Sicilia
Carmelo Nicotra
Stazione Consorziale Sperimentale di Granicoltura per la Sicilia
Sebastiano Blangiforti
Stazione Consorziale Sperimentale di Granicoltura per la Sicilia
Alfio Spina
CREA Research Centre for Cereal and Industrial Crops
Salvatore Cosentino
University of Catania
Angela Roberta Lo Piero (rlopiero@unict.it)
University of Catania

Research Article

Keywords: Durum wheat, River wheat, SSR marker, landrace, genetic structure, agrobiodiversity

Posted Date: May 17th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-504806/v1

Version of Record: A version of this preprint was published at Plants on January 4th, 2022. See the published version at https://doi.org/10.3390/plants11010130.

Abstract

Cereal landraces are a very valuable resource in contemporary agriculture. A renewed focus for breeding purposes could ameliorate some negative consequences of modern agriculture and conventional breeding, such as the loss of genetic diversity. A strategy combining molecular genotyping and characterization of morpho-agronomic traits related to productivity is proposed to assess a group of tetraploid wheat landraces named “Bufala”, historically cultivated in the mountain areas of Sicily and used for the production of a traditional bread type locally appreciated. A total of 55 SSR molecular markers were used to detect pattern of diversity for 30 rivet and durum wheat genotypes. Furthermore, phenotyping was than conducted for 8 morpho-agronomic traits. Discriminant analysis of principal components (DAPC), STRUCTURE and phylogenetical analysis allowed to identify three groups, two of them genetically close and including both “Bufala” and “Bufala-related” rivet landraces. To the third group old and more recent durum wheat varieties, constituting the outgroup, were assigned. Clustering was confirmed by Principal Component Analysis (PCA). Finally, a correlation analysis showed as “Bufala” genotypes are characterized by lower ear density compared with the other studied genotypes, a major ear length and a later earing time. The generated knowledge about the levels of diversity and population structure could be an important contribution for parent selection in tetraploid wheat breeding programs in order to define a plant ideotype suitable for low-input and organic cropping systems, as well as for germplasm conservation and management.

1. Introduction

Wheat landraces are a very valuable genetic resource for the different contemporary cereal-based farming systems. Camacho Villa and collaborators proposed the following definition: “a landrace is a dynamic population of a cultivated plant that has historical origin, distinct identity and lacks formal crop improvement, as well as often being genetically diverse, locally adapted and associated with traditional farming systems” (Villa et al. 2005). Until the end of the nineteenth century landraces were the principal resource of agricultural production (Harlan 1975). A gradual replacement during the early decades of the 20th century by selected component pure lines and modern cultivars happened. Nevertheless, the persistence of landraces in different environments was due to their increased stability, accomplished through generations of natural and deliberate selection for valuable genes for resistance to biotic and abiotic stresses and intergenotypic competition, as well as to their favorable morpho-physiological and agronomic traits tightly linked with the high yield ability and quality of products. Besides, landraces represent the maximum expression of the adaptation to the specific environmental conditions and agronomic management of the areas were they evolved (Moragues et al. 2007). There are a number of wild species, landraces, and traditional cultivars within the *Triticum* genus that constitute the wheats of the world. The main center of diversity is the Fertile Crescent, extending from the Mediterranean coast in the west to the east. In this region, diploid and polyploid *Triticum* species coexist in mixed populations and exhibit a wide range of morphological and ecological diversity (Hoisington et al. 1999). Among polyploid species, tetraploid wheats (*Triticum turgidum* L.) belong to a taxonomic category that includes

genetically and morphologically different entity, and their evolution under domestication has not been fully explained. Archaeological findings and genetic studies indicate that emmer [*Triticum turgidum* L. subsp. *dicoccum* (Schrank) Thell.], the first domesticated form of tetraploid hulled wheat originated from the tetraploid wild ancestor in the western half of the Fertile Crescent 7,800-7,500 before present. Free-threshing tetraploid naked wheats, rivet (*Triticum turgidum* L. subsp. *turgidum*) also called poulard, cone or english wheat, and afterwards durum [*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.], evolved from emmer in the Near East and spread through the north side of the Mediterranean area reaching the Iberian Peninsula and Algeria from Italy. It has also been hypothesized that the rivet wheat arrived in England with the Normans from Sicily, and according to Percival (Percival 1921) surely there are a group of rivet landraces adapted only for growing in countries bordering the Mediterranean. Through the evolution in mountain environments, rivet wheat has acquired rusticity in terms of could tolerance, ability to grow in marginal soils, weed competitiveness and resistance to diseases. Moreover, due to the kernel hardness not fully vitreousness as durum wheat, the grain of rivet wheat has been traditionally used preferably for bread making or for homemade pasta and biscuit preparation. However, nowadays, durum wheat remains the most important cultivated tetraploid wheat in the Mediterranean basin, Canada and USA. In the Mediterranean region, it is mainly grown in the Middle and Near East regions and North Africa. Based on cytological and molecular analyses, *Triticum turgidum* is believed to be originated from a hybridization between *Triticum urartu* (AA genome, n = 7) and an unknown diploid species (BB genome, n = 7) closely related to *Aegilops speltoides* (Dubcovsky and Dvorak 2007). Thus, tetraploid wheats as rivet and durum wheat, are a self-pollinated allotetraploid cereal (harboring two genomes with genomic formula: AABB) with 28 chromosomes (2n = 4x = 28). In addition to the taxonomical classifications proposed by both Van Slageren and MacKey, which consider rivet and durum wheats as subspecies of *T. turgidum*, MacKey further classified these forms as covarieties of the subspecies *T. turgidum* subsp. *turgidum* (*T. turgidum* subsp. *turgidum* conv. *durum*), whereas Dorofeev and colleagues considered tetraploid wheats as individual species *T. durum*. Nevertheless, the more recent results obtained by Oliveira and collaborators (Oliveira et al. 2012) on the basis different marker types (nuSSRs, ISBPs, cp-haplotypes) and several analysis methods (STRUCTURE, PCA, genetic distance approaches), support MacKey's taxonomical classification, which consider rivet and durum wheats as varieties of the same taxon. The authors also suggest that rivet landraces constitute an easily transferable source of genetic variation for durum wheat breeding programs. Fortunately, although rivet has been neglected and it disappeared from cultivation during the last century, accessions have been preserved by the inclusion in germplasm bank collections and are available for new breeding activity. The knowledge of the extent and pattern of genetic diversity within and among wheat populations is a key factor for the identification of useful genotypes for plant breeding purposes and to better understand the crop requirements to design appropriate collection and conservation strategies (Asmamaw et al. 2019). Furthermore, a renewed focus on wheat landraces for breeding purposes could relieve some negative consequences of intensive agriculture and conventional breeding, such as the irrational and/or excessive use of auxiliary input, excessive homogeneity of cropping systems, the loss of genetic diversity (Tilman 1998), and the stagnation of yields in marginal cereal areas (Annicchiarico and Pecetti 1998). This is also functional for the definition of a plant ideotype suitable for low-input farming systems, mainly smallholder and organic

(Anastasi et al.). In Italy, tetraploid wheats, especially durum wheat, have a long-time tradition of growing and breeding, and accessions collected in Southern Italy, which include rivet germplasm, now preserved ex situ, are a valuable genetic resource. A large number of Italian and some North African landraces of tetraploid wheats were grown in Italy since the end 1800 and after 1920 were gradually substituted by the pure line varieties selected from Italian landraces, but also from Syria-Palestinia and North Africa (Lybia) germplasm. Starting from the second half of the last century, also the internal hilly areas of Sicily characterized by the cultivation of wheat local populations, including some exaploid wheat (*Triticum aestivum aestivum*), have drastically contracted until almost total disappearance due to the advent of an increasing number of novel improved varieties, which are able to exploit grater levels of auxiliary input, particularly for nitrogen nutrition and weed control. In the mountain environments, this phenomenon occurred much more slowly because the so-called “modern” varieties hardly adapted to the extreme variability of local pedoclimatic conditions. The evolution of these landraces has been characterized not only by the particular pedoclimatic conditions of the Sicilian mountain areas, but also from the socio-economic and productive needs that this territory imposes. The selection exercised by the farmer-breeder on these local populations over the centuries has mainly taken into account two fundamental aspects: the production of semolina, suitable for processing into large-sized traditional breads (from 2 to 4 kg) suitable for long-term shelf-life during transhumance, and the high production of straw essential in cattle farms. The considerable advances in molecular genotyping and databasing technologies in recent years are beginning to make the variation and resources of landraces more accessible for exploitation. High-throughput genotyping enables Genbank accessions with uncertain provenance to be elucidated and thereby enables validation of associated phenotypic data, making them much more useful (Newton et al. 2009). Molecular markers, such as RFLP, SSR and SNP have been successfully used for identification of cultivars, diversity estimates, and genetic relationships assessment in crops including rivet and durum wheat (Autrique et al. 1996; Maccaferri et al. 2003; Kabbaj et al. 2017; Fiore et al. 2019). For their high polymorphism, codominance and locus specificity, simple sequence repeats or microsatellite (SSRs) markers have proved to be highly efficient molecular tools for the characterization of durum wheat germplasm collections (Maccaferri et al. 2003; Moragues et al. 2007; Figliuolo et al. 2007; Ruiz et al. 2012; Laidò et al. 2013; Sahri et al. 2014). Different authors in the past years developed physical consensus maps of SSR markers in both soft wheat and durum wheat chromosomes (Röder et al. 1998; Somers et al. 2004; Paux et al. 2012). To date most studies on Italian durum germplasm have analyzed collections including old and new elite varieties for morphophysiological and qualitative traits (De Vita et al. 2007; Motzo and Giunta 2007), and the use of molecular markers was focused on temporal trends of diversity (Hoisington et al. 1999; Giunta et al. 2007; Kabbaj et al. 2017), relatedness among genotypes (Maccaferri et al. 2003), genetic structure (Martos et al. 2005), also in comparison to *Triticum turgidum* L. subspecies (Laidò et al. 2013). In a recent study (Kabbaj et al. 2017) a panel of 370 durum wheat genotypes including 35 Italian genotypes were genotyped using 500 single nucleotide polymorphism (SNPs) markers. In 2018, Marzario and colleagues (Marzario et al. 2018) used a smaller number (44) of simple sequence repeats (SSR) molecular markers to detect pattern of diversity for 136 accessions collected in the South Italy over time, to identify the genepool of origin, and establish similarities with 28 Italian varieties with known pedigree grown in Italy over the same time-period. They also conducted

phenotyping for 12 morphophysiological traits of agronomic interest thus obtaining enough information on the genetic structure of durum wheat genotypes for a quick screening of the germplasm collection. More recently, in 2019, Asmamaw and collaborators (Asmamaw et al. 2019) assessed the magnitude and pattern of genetic diversity in Ethiopian durum wheat landraces by SSR molecular markers analysis. Furthermore, Fiore and collaborators in 2019 (Fiore et al. 2019) characterized a collection of durum wheat landraces from Sicily, using single nucleotide polymorphisms (SNP) markers, together with agro-morphological, phenological and quality-related traits. For the above-mentioned reasons, in this study a strategy combining molecular genotyping and morpho-agronomic traits is proposed to characterize a group of rivet wheats named “Bufala”, historically cultivated in the mountain areas of Sicily (from 800 to 1,200 m a.s.l.) and used for the production of a traditional bread type locally appreciated. In particular, a total of 55 SSR molecular markers were utilized to detect pattern of diversity for 30 tetraploid wheat genotypes, a collection of twenty “Bufala” and seven “Bufala-related” rivet landraces, in comparison with an outgroup of three improved durum wheat varieties., Furthermore, phenotyping was than conducted for a set of significant morpho-agronomic traits, potentially useful for breeding purposes to obtain new genotypes suitable for low-input farming systems.

2. Materials And Methods

2.1. Plant material and DNA extraction

A total of 30 Sicilian tetraploid wheat genotypes [(*Triticum turgidum* L. ssp. *turgidum* convar. *durum* (Desf.), 2n = 4x = 28; genomes AABB] including 27 rivet wheat landraces belonging to a “Bufala” and a “Bufala-related” (genotypes genetically close to “Bufala” germplasm) groups and an outgroup of three improved varieties of durum wheat, two old (“Bidì03” and “Capeiti”) and a more recent one (“Simeto”), which have in common the old variety released in 1915 “Senatore Cappelli” in their pedigree, were chosen for this study. The list of the accessions and their origin is reported in Online resource 1 - Table S1. The grains were provided by “Stazione Consorziale Sperimentale di Granicoltura per la Sicilia (Santo Pietro, Caltagirone, Catania, Italy), and come from tetraploid wheats grown in a field trial laid-out according to a randomized-block design replicated three times conducted during the 2018–2019 growing season in the experimental station sited in Vaccarizzo (Lat. 37,119000° - Lon. 14,521000° – 316 m asl) (Catania), adopting a low input agronomic management consisting in 30 kg ha-1 N supply at sowing and no-chemical weed control during the cropping cycle. Grains were sampled from a collection of ten spikes from a representative group of plant of each accession. Kernels were soaked in sodium hypochlorite solution 1% for 15 minutes and then rinsed with sterile water for three minutes in a laminar flow hood. Ten sterile kernels were then placed in petri dishes on filter paper moistened with sterile water in dark conditions until germination. Once germinated the epicotyls were frozen with liquid nitrogen and grounded with mortar and pestle for the following genomic DNA extraction. DNeasy® Plant Mini Kit (Qiagen, Hilden, Germany) was used for the DNA extraction and the purity and concentration of DNA were determined spectrophotometrically at 260 and 280 nm by using Nano Drop® Spectrophotometer ND 1000 (NanoDrop Technologies, Wilmington, DE, USA).

2.2. Molecular characterization

In order to estimate the genetic diversity, the genotypes of tetraploid wheats were characterized using 55 SSR markers. All markers were selected based on the chromosomal position (chromosomes 1 to 7 A and 1 to 7 B) and on the Polymorphic Information Content (PIC) reported in the previous studies, in order to give uniform coverage of tetraploid wheat chromosomes and high informativity. Detailed information on the used markers such as the chromosomal position, the sequence of the primers used, the temperature of annealing, the repeated motifs and the bibliographic sources are listed in Online resource 1 - Table S2.

2.3. Markers amplification and sequencing

PCR amplification of genomic DNA was performed with a final volume of 15 µL containing 50 ng of genomic DNA, 0.5 mM of each dNTP, 0.25 µM of each primer, and 0.75 U of HSTaq polymerase as follows: 5 min at 95°C, 30 cycles with 30 seconds at 95°C, 30 seconds at either 55 or 60°C (depending on the locus) and 30 seconds at 72 ◦C, followed by a final extension step of 30 min at 60°C. PCR products were sequenced at CD Genomics (Shirley, NY, USA) using 3730XL Sequencer (Applied Biosystems, Foster City, CA, USA): each well of a 96-well plate was added with 9 µl of a molecular weight internal standard and a mixture of formamide (0.5:8.5), and 1.0 µl PCR product. The detected raw data ".fsa" file was imported into analysis software GENEMAPPER v3.2 (Applied Biosystems, Foster City, CA, USA) for analysis.

2.4. Genotyping and statistical analysis

Data from sequencing were used to calculate statistical parameters such as number of alleles (Na), observed heterozygosity (Ho), expected heterozygosity (He) and polymorphism information content (PIC) of each SSR locus. The software used to obtain these parameters was Cervus 3.0.7 (Field Genetics Ltd., London, UK). Private alleles were also calculated using GenAlEx 6.503, a package for population genetic analysis that runs within Microsoft Excel (Peakall and Smouse 2012). Population structure of genotypes was examined by first applying the Discriminant analysis of principal components (DAPC) (Jombart et al. 2010), a multivariate method designed to identify and describe clusters of genetically related individuals. Discriminant analysis of principal components was performed using the adegenet package (Jombart 2008) in R (https://www.rstudio.com/products/team/. Accessed 13 Apr 2021). The optimal number of clusters was determined using the *find.clusters* function which implements successive K-means clustering. The rate of decrease of the Bayesian Information Criterion (BIC) was examined, and the number of clusters was determined as the value of K above which BIC values decreased or increased only subtly. The *dapc* function was then applied to describe the relationship between the inferred groups. In order to obtain reliable group membership probabilities and to avoid overfitting, we retained only the three first principal components (PCs) from the preliminary data transformation step (indicated to be the optimal number based on the *optim.a.score* function). A model-based approach was further applied. This approach was implemented with STRUCTURE software (FALUSH et al. 2007). At first, the number of subgroups (Cluster = K) was set from 1 to 10. Ten independent simulations were performed for each K setting using the admixture model, with each simulation set to a 5000 burn-in period and 50,000 Markov

chain Monte Carlo (MCMC) repetitions. The optimal number of K was then determined by using STRUCTURE HARVESTER (Earl and vonHoldt 2012), with which the Delta K statistical test was calculated, combined with the probability that each preset K is the correct one. Finally, the number of K with the highest Delta K value was chosen (Online resource 1 - Figure S1).

2.5. Genetic distance

Once the genetic structure of the genotypes had been obtained, the examination of their degree of differentiation was performed. DARwin 6.0 software was used to calculate the Nei's unbiased genetic distance. Through the neighbor joining method (NJ) (Saitou and Nei 1987), the dendrogram was built by comparing single genotypes, setting 1000 as the bootstrap value.

2.6. Morpho-agronomic characterization

During the growing cycle, a sample of ten plants per field replicates of each studied genotypes were subjected to measurements of morpho-agronomic traits related to crop productivity, following also the Zadoks scale system, and after the harvest on a sample of ten representative ears. The evaluated traits, scored according to descriptors for wheat defined by the CPVO protocol (CPVO-TP/120/3), are reported in Online resource 1 - Table S3: *Habitus* (erect, prostrate), time of earing (early, late), culm height (cm), ear length (short, long), awns length (cm), ear shape (thin, thick), ear density (lax, dense), 1000-kernel weight (g). Principal component analysis (PCA) was performed using R, in order to determine the overall morphological traits distinctiveness, and to investigate the relationships between them. Finally, the Pearson correlation coefficient ($p < 0.05$) was also calculated using the *round(cor)* R function and a scatter plot with the correlation coefficients was developed with the R/corrplot package (https://cran.r-project.org/web/packages/Corrplot/index.html).

3. Results

3.1. Genetic profile

Samples were divided into three groups on the basis of the germplasm type: "Bufala", "Bufala-related" (genotypes genetically close to "Bufala" germplasm) rivet wheat landraces, "Outgroup" of improved durum wheat varieties (two old varieties "Bidì03" and "Capeiti" and a more recent one "Simeto"). All SSR markers showed polymorphism (PIC > 0) and a total of 384 alleles were detected across the 30 genotypes (Table 1). The average number of alleles (Na) per SSR was 6.98 (Table 1), ranging from one allele (Xgwm415) to 22 alleles (Xgwm268 and Xgwm369) (Online resource 1 - Table S4). The total number of alleles per locus is reported in Online resource 1 - Table S4. Furthermore, expected heterozygosity (H_e) across the total genotypes was 0.60 (Table 1), and ranged from 0.18 (Xgpw2239) to 0.95 (Xgwm268 and Xgwm369), while the observed average heterozygosity (H_o) was 0.34 (Table 1), with a minimum of 0 and a maximum of 1 (Online resource 1 - Table S4). When considering the germplasm groups, the "Bufala" genotypes highlight a higher average number of alleles (5.56) and a higher PIC (0.50) when compared with the other groups (Table 1). The subsequent genetic analysis were performed by using the most

informative SSR markers, considering PIC ≥ 0.44 as threshold, since a lower PIC is considered scarcely informative (Hildebrand et al. 1992).

Table 1
Genetic diversity estimated for tetraploid wheats. Na = average number of alleles; H_o= average observed heterozygosity; H_e= average expected heterozygosity; PIC = average polymorphic information content.

Genotype group	N° of samples	Na	H_o	H_e	PIC
Bufala	20	5.56	0.35	0.55	0.50
Bufala-related	7	4.16	0.34	0.55	0.48
Outgroup	3	2.25	0.27	0.42	0.32
Total	30	6.98	0.34	0.60	0.56

A list of private alleles, alleles that are found only in a single population among the broader collection, of each genotype is also reported in Fig. 1. The figure shows that Bufala Bianca 04 (BB-04), Ciciredda 03 (CIC-03) and Bivona 04 (BIV-04) are the three landraces with the highest number of private alleles.

3.2. Genetic structure

The BIC analysis used to assess the optimal number of clusters identified three genetic clusters (K = 3). A scatterplot of the first two principal components of the DAPC, accounting for 16% of the total variance, is shown in Fig. 2 in order to describe the relationship among the clusters. The distribution of the genotypes among the three clusters is reported in Table 2.

Table 2
Distribution of genotypes in three clusters obtained by discriminant analysis of principal components (DAPC).

Genotype group	DAPC		
	C1	C2	C3
Bufala	5	13	2
Bufala-related	3	4	0
Outgroup	0	0	3

Cluster 1 and 2 (C1 and C2) resulted to be genetically similar and partially overlapped. In these two clusters are distributed both “Bufala” and “Bufala related” genotypes. Thirteen out of twenty “Bufala” landraces (65%) were grouped in C2, thus representing the main “Bufala” cluster, together with four “Bufala-related” landraces (BIV-03, BIV-04, PAO-01 and PAO-02). Five “Bufala” landraces (BRL-01, BT-01,

BS-02, BB-04, BG-03) and the three Ciciredda landraces (CIC-01, CIC-02 and CIC-03) were grouped in C3, suggesting a genetic differentiation from the “Bufale” core cluster. Cluster 3 (C3) is clearly separated from the other clusters; it includes the “outgroup” of durum wheat varieties (BIDI-03, SIM and CAP-8) and two “Bufala” landraces (BRCa-01 and BRCb-01). Considered the low discriminant power of the DAPC analysis and the partial overlapping when considering C2 and C3 clusters, a STRUCTURE analysis was further performed. The analysis consists in a Bayesian model-based clustering method. The number of subpopulations (K) was identified based on Delta K values (EVANNO et al. 2005). The highest value of Delta K was found at three clusters (K = 3) (Online resource 1 - Figure S1). The STRUCTURE bar graphic also provides information on the level of admixture in the study sample. At K = 3, 29 genotypes out of 30 (97%) were assigned to one or another group with more than 70% posterior membership probability (Fig. 3). The remaining “Bufala rossa lunga 01” (BRL-01) resulted with 50% probability belonging to K1 and 50% to K2. Individual assignments provided by STRUCTURE resulted more discriminant from those provided by DAPC.

As shown in Fig. 3, the three clusters were clearly separated: the first cluster (K1; green color) was composed of 7 landraces of which 4 belonging to “Bufala” and 3 belonging to “Bufala-related” group. Here the three “Bufala bianca” (BB-02, BB-03, BB-04) clustered together with “Paola” (PAO-01, PAO-02), “Bivona 03” (BIV-03) and “Bufala rossa lunga 03” (BRL-03). The second cluster (K2, blue color) included 13 “Bufala” landraces, “Ciciredda” (CIC-01, CIC-02, CIC-03) and “Bivona 04” (BIV-04). Finally, the third cluster (K3; red color) included the “outgroup” of durum wheat varieties and two “Bufala” rivet landraces (BRCa-01 and BRCb-01).

3.3. Phylogenetic analysis

A phylogenetic analysis was also carried out. The analysis based on Nei (Nei 1978) genetic coefficient and the neighbor joining algorithm, generated a dendrogram underlining three main groups overlapping with the clusters identified by STRUCTURE analysis (Fig. 4). Bootstrap higher than 70% resulted in the most important nodes, avoiding any misclassifications (Fig. 4). As expected, the “outgroup” of durum wheat varieties together with BRCa-01 and BRCb-01 rivet landraces showed the highest values of genetic distance from the other groups.

3.4. Morpho-agronomic traits

The Principal Component Analysis (PCA) performed on a set of morpho-agronomic traits related to productivity showed some differences among phylogenetic groups (Fig. 5). The first two components explained 47% of the total variance. The PC1 allowed to discriminate the three groups identified by the phylogenetic analysis, and separated the genotypes by time to earing stage (E), ear length (EL) and ear density (ED) (Fig. 5).

Indeed genotypes belonging to Group III (BIDI-03, SIM, CAP-8, BRCb-01 and BRCa-01) are associated with a higher ear density, an earlier earing time and lower ear length. An opposite condition was noted in the genotypes belonging to Group I (Paola, Bufala Bianca, Bivona and Bufala rossa lunga). Group II is

characterized by a high variability of traits such as *habitus* (HA) and ear shape (ES), which are the traits described by the PC2. These evidences were confirmed by Pearson correlations (Fig. 6). Figure 6a shows the significant correlation indexes (pvalue < 0.05) of all the genotypes considering also the group membership. The analysis confirms that there is a negative correlation between ear length (EL) and group membership (GR) (from Group I to Group III) (-0.78) and between time of earing (E) and group membership (GR) (-0.6) and a positive correlation between ear density (ED) and group membership (GR) (0.64). A negative correlation between ear length (EL) and ear density (ED) (-0.71) was also found, indicating that as expected an increased ear length is associated with a lower seed density. When considering the trait correlations within groups, in Fig. 6 (b, c and d, referred respectively to Group I, II and III) is indicated a high positive correlation between *habitus* (HA) and time of earing (E) (Fig. 6b), a positive correlation between ear length (EL) and ear shape (ES). Whereas a negative correlation between ear length (EL) and awn length (AL) and between awn length (AL) and ear shape (ES) (Fig. 6c) were also found. Finally, in Group III there is a strong inverse correlation between *habitus* (HA) and ear shape (ES) (Fig. 6d). All the described correlations are significant (p-value < 0.05).

4. Discussion

In Italy tetraploid wheats, particularly durum wheat, have a long-time tradition of growing and breeding, and germplasm of local populations collected *in situ* and *ex situ* in Southern Italy, represents a valuable resource to preserve the cereal genetic diversity ensuring food security in the future. For this reason, the identification of precious traits, diversity estimates, and genetic relationship assessments are essential in order to take a rational advantage from these wheat landraces. Genetic selection decreases the variability over time as highlighted by Marzario and co-workers (Marzario et al. 2018). In fact, they found that the amount of genetic diversity decreased in 22 accessions collected from 1983 to 2003, when most obsolete varieties had been already replaced, making old varieties and landraces a precious source of diversity (Marzario et al. 2018). A group of tetraploid rivet wheats named “Bufala”, historically cultivated in mountain areas of Sicily and used for a particular type of bread production and other locally appreciated bakery products, were evaluated in this study by combining molecular genotyping and morpho-agronomic characterization. Different studies led to the construction of high resolution microsatellite maps for both soft and durum wheat covering the seven homoeologous chromosome groups (Röder et al. 1998; Somers et al. 2004; Marone et al. 2012). The availability of such maps allowed researchers to characterize, genotypes and compare *Triticum* species, resulting in a genetic relationship estimation among genetically, temporally and geographically distant varieties and accessions, and obtaining useful information for breeding purposes (Maccaferri et al. 2003; Laidò et al. 2013; Abbasov et al. 2018; Asmamaw et al. 2019). For example in 2018 and 2019 a first classification and evaluation of Sicilian old germplasm was obtained by using both SSR and SNP markers (Marzario et al. 2018; Fiore et al. 2019). Here a total of 55 SSR molecular markers where sequenced and analyzed on 30 Sicilian genotypes of tetraploid wheats, 20 of which “Bufala” rivet landraces, 7 “Bufala-related” rivet landraces, in comparison with 3 improved varieties of durum wheat, two old and a more recent one, which have in common the old variety “Senatore Cappelli” in their pedigree (“outgroup”). The genetic diversity estimation obtained was in

line with what observed in other similar studies (H_e = 0.60) (Maccaferri et al. 2003; Marzario et al. 2018), resulting in an higher H_e (0.55) of the “Bufala” rivet landraces compared with the improved varieties group (H_e=0.42). Thirty-six loci with an average PIC ≥ 0.44 where then selected for the subsequent analysis in order to have a good discriminant power. Among the germplasm under study, Bufala Bianca 04 (BB-04), Ciciredda 03 (CIC-03) and Bivona 04 (BIV-04) evidenced the highest number of private alleles, making these landraces eligible genetic resource for characterization and genotypes traceability. Discriminant Analysis of Principal Components (DAPC) and STRUCTURE analysis, providing information about genetic structure, both grouped the studied genotypes in three different clusters. The same approach was carried out by Marzario and co-workers (Marzario et al. 2018), and both the methods resulted very informative and complementary in obtaining the genetic structure of the tetraploid wheat germplasm collection coming from Sicily compared with other Italian accessions. Based on discriminant analysis of principal components (DAPC) and STRUCTURE analysis they identified six groups, and the assignment of varieties reflected the genetic basis and breeding strategies involved in their development. In this work, although both the analysis clustered the “outgroup” durum wheat varieties in a well-defined cluster (C3 for DAPC and K3 for STRUCTURE), STRUCTURE turned out to be better in discriminating genetically related rivet wheat landraces such as those of “Bufala” and “Bufala-related” groups. STRUCTURE analysis allowed Laidò and collaborators to distinguish durum wheat cultivars from the other tetraploid subspecies, and two distinct subgroups were also detected within the tetraploid wheat subspecies, which is in agreement with their origin and year of release (Laidò et al. 2013). Cultivars belonging to the aforementioned groups were distributed between C1-C2 (Fig. 2) and K1-K2 (Fig. 3). By DAPC analysis C1 and C2 partially overlap leading to an ambiguous classification of different genotypes, whereas STRUCTURE resulted in a significant cluster allocation of 29 genotypes out of 30: only Bufala Rossa lunga 01 (BRL-01) had a posterior probability < 0.70. The phylogenetic analysis confirmed the results obtained by STRUCTURE both in terms of number of cluster membership and genotype assignment, giving a further demonstration of its goodness with respect of DAPC analysis in discriminating genetically-related genotypes. It is interesting to notice that the Bufala bianca (BB) landrace clustered together with Paola (PAO), Bivona (BIV) and Bufala Rossa Lunga (BRL) (K1 in Fig. 3, Group I in Fig. 4), probably indicating a common origin. The same occurred in the case of Ciciredda (CIC), Bufala Nera (BNL and BNC) and other Bufala landraces such as Flascio (BF), Salice (BS), Troina (BT), Cerami (BC) e Gangi (BG) that clustered in K2 (Group II in Fig. 4), indicating a different origin or genetic differentiation process. As expected, the improved varieties Bidì 03 (BIDI-03), Simeto (SIM) and Capeiti (CAP-8), used as outgroup, clustered together in K3 (Fig. 3) and Group III (Fig. 4), resulting in a higher genetic distance. Surprisingly, two Bufala Rossa Corta genotypes (BRCa-01 and BRCb-01) clustered together with the improved varieties, probably representing the point from they were originated. This result is in accordance with those obtained by the SNP characterization of Fiore and collaborators demonstrating that BRCb01 clustered together with Simeto (SIM) and Bidì03 (BIDI03) (Fiore et al. 2019). Nevertheless, our findings partially support those of Oliveira and colleagues (Oliveira et al. 2012), which in all their marker systems found strong genetic similarity between rivet and durum wheats and concluded that probably the two subspecies were originated from a common domesticated ancestor. The Authors, however, also affirmed that the adaptation of plants to specific conditions after the species was introduced into Europe could have favored the evolution of

landraces with distinct morphological characteristics, such as the dis-tinct ear form in rivet and the cold tolerance in comparison with durum wheat. These differences would be maintained by artificial selection giving rise to the agronomically distinct rivet and durum wheat, but the selective pressure would not have been so strong to create a distinct genetic pool between the two tetraploid wheats. They also speculated that the distinction between rivet and durum wheats is simply an artefact based on the criteria used by early botanists, emphasizing the differences rather than similarities between groups, whereas traditional farmers might have simply thought in terms of varieties with similar agronomic properties, like thresh ability, considering that all cultivated tetraploid wheats are inter-fertile with one another. Additional information was inferred by the genetic assessment combined with a morpho-agronomic characterization of the Sicilian tetraploid wheats. The Principal Component Analysis (PCA) performed on a set of representative traits related to productivity, also identified three distinct groups, totally overlapping with those identified by the genetic analysis. Also considering the morphological traits, BRCa-01 and BRCb-01 rivet landraces are more similar to the durum wheat improved variety BD-03 and SIM and to CAP-8 in terms of time of earing, ear density and ear length. On the other hand, the higher variability was found within “Bufala” and “Bufala-related” genotypes for the *habitus* of plants, awns length and ear cross shape. These evidences were confirmed by Pearson correlations, showing a positive correlation ($p < 0.05$) between *habitus* and time of earing (0.94) and between ear length and ear cross shape (0.76) in “Bufala” and Bufala-related” wheats. Furthermore, improved cultivar resulted to have a smaller (ear length *vs* group = -0.78), but a more ear density (ear density *vs* group = 0.64).

5. Conclusions

Wheat landraces are a very precious genetic resource in view of a transition of cereal-based farming systems towards greater sustainability. A renewed focus on this germplasm for a new breeding approach could mitigate certain negative consequences of intensive agriculture and conventional breeding, such as the excessive use of chemical input, loss of genetic diversity due to the high crop homogeneity based on monoculture farming and the limited yield increase in marginal cereal areas. For their high polymorphism, codominance and locus specificity, simple sequence repeats or microsatellite (SSRs) markers have proved to be highly efficient molecular tools for detecting genetic variation and characterizing germplasm collections. In wheat, SSRs usually give a unique fragment specific to each homoeologous copy and show a high level of polymorphism compared with other types of molecular markers including SNPs (Paux et al. 2012). Moreover, they frequently reveal a higher number of alleles at each locus making them very effective to study genetic relationships (Roussel et al. 2005). In fact, SSR markers allowed to identify a number of unique alleles, specific for almost all the landraces analyzed, representing a relevant information considering the similarity within groups. Both the morpho-agronomic traits and SSR markers used in the present study were equally appropriate to provide a first overview of the genetic diversity levels and of the population structure within the rich tetraploid wheat germplasm collection available in Southern Italy. SSRs appeared powerful for evaluating genetic diversity and classifying different rivet wheat landraces, due to their reproducibility and informativity. Moreover, selected SSR markers were also able to discriminate landraces from improved varieties and within populations and being to such extent

promising their use suggests to enlarge the germplasm collection to analyze in the future. The investigation of population structure suggested the genetic potential of landraces for the detection of unexplored sources of variation and allowed to identify groups of accessions differentiated both at molecular level and for morpho-agronomic traits. Thus, the SSR panel will allow to organize an efficient system for genetic traceability of wheat. The generated knowledge about the levels of diversity and population structure could be an important contribution for parent selection in tetraploid wheat, especially in more widespread durum wheat, breeding programs for germplasm conservation and management in order to provide varietal innovation to support low-input and polyculture farming.

Declarations

Statement: The use of plants or plant parts in the present study complies with international, national, and institutional guidelines

Funding: or “This research was funded by NAME OF FUNDER, grant number XXX” and “The APC was funded by XXX”. Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding. Any errors may affect your future funding.

Conflicts of Interest: The authors declare no conflict of interest.

Supplementary information: Online resource 1, Table S1: list of accessions and their origin; Table S2: SSR markers detailed information; Table S3: morphological descriptors and their scores (UPVO/CPVO, 2011); Table S4: genetic diversity of each locus; Figure S1: STRUCTURE estimation of the number of subpopulations for K ranging from 1 to 10 by Delta K values (ΔK) (B).

Author Contributions: Conceptualization, A.R.L.P., A.S. (Angelo Sicilia) and U.A.; methodology, A.R.L.P., A.S. (Angelo Sicilia) and A.S. (Alfio Spina); software, A.S. (Angelo Sicilia); formal analysis, A.S. (Angelo Sicilia); investigation, A.S. (Angelo Sicilia), M.B. and S.M.; data curation, A.R.L.P., A.S. (Angelo Sicilia), S.B. and U.A.; writing—original draft preparation, A.S. (Angelo Sicilia); writing—review and editing, A.R.L.P., S.B., U.A. and S.L.C.; visualization, S.B., U.A and C.N.; supervision, A.R.L.P.; funding acquisition, C.N., U.A. and S.L.C. All authors have read and agreed to the published version of the manuscript.

References

1. Abbasov M, Akparov Z, Gross T, et al (2018) Genetic relationship of diploid wheat (*Triticum* spp.) species assessed by SSR markers. Genet Resour Crop Evol 65:1441–1453. https://doi.org/10.1007/s10722-018-0629-2
2. Anastasi U, Corinzia SA, Cosentino SL, Scordia D Performances of Durum Wheat Varieties Under Conventional and No-Chemical Input Management Systems in a Semiarid Mediterranean Environment. https://doi.org/10.3390/agronomy9120788

3. Annicchiarico P, Pecetti L (1998) Yield vs. morphophysiological trait-based criteria for selection of durum wheat in a semi-arid Mediterranean region (northern Syria). F Crop Res 59:163–173. https://doi.org/10.1016/S0378-4290(98)00116-6
4. Asmamaw M, Keneni G, Tesfaye K (2019) Genetic Diversity of Ethiopian Durum Wheat (*Triticum durum* Desf.) Landrace Collections as Reveled by SSR Markers. Adv Crop Sci Technol 07: https://doi.org/10.4172/2329-8863.1000413
5. Autrique E, Nachit MM, Monneveux P, et al (1996) Genetic Diversity in Durum Wheat Based on RFLPs, Morphophysiological Traits, and Coefficient of Parentage. Crop Sci 36:735–742. https://doi.org/10.2135/cropsci1996.0011183X003600030036x
6. De Vita P, Nicosia OLD, Nigro F, et al (2007) Breeding progress in morpho-physiological, agronomical and qualitative traits of durum wheat cultivars released in Italy during the 20th century. Eur J Agron 26:39–53. https://doi.org/10.1016/j.eja.2006.08.009
7. Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316:1862–1866
8. Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359–361. https://doi.org/10.1007/s12686-011-9548-7
9. Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. Mol Ecol 14:2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x
10. Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol Ecol Notes 7:574–578. https://doi.org/10.1111/j.1471-8286.2007.01758.x
11. Figliuolo G, Mazzeo M, Greco I (2007) Temporal variation of diversity in Italian durum wheat germplasm. Genet Resour Crop Evol 54:615–626. https://doi.org/10.1007/s10722-006-0019-z
12. Fiore, Mercati, Spina, et al (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
13. Giunta F, Motzo R, Pruneddu G (2007) Trends since 1900 in the yield potential of Italian-bred durum wheat cultivars. Eur J Agron 27:12–24. https://doi.org/10.1016/j.eja.2007.01.009
14. Harlan JR (1975) Our vanishing genetic resources. Science 188:618–621. https://doi.org/10.1126/science.188.4188.617
15. Hildebrand CE, Torney DC, Wagner RP (1992) Informativeness of Polymorphic DNA Markers. Los Alamos Science 20:100-102
16. Hoisington D, Khairallah M, Reeves T, et al (1999) Plant genetic resources: What can they contribute toward increased crop productivity? Proc Natl Acad Sci U S A 96:5937–5943. https://doi.org/10.1073/pnas.96.11.5937

17. Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. https://doi.org/10.1093/bioinformatics/btn129
18. Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. BMC Genet 11:94. https://doi.org/10.1186/1471-2156-11-94
19. Kabbaj H, Sall AT, Al-Abdallat A, et al (2017) Genetic Diversity within a Global Panel of Durum Wheat (*Triticum durum*) Landraces and Modern Germplasm Reveals the History of Alleles Exchange. Front Plant Sci 8:1277. https://doi.org/10.3389/fpls.2017.01277
20. Laidò G, Mangini G, Taranto F, et al (2013) Genetic Diversity and Population Structure of Tetraploid Wheats (*Triticum turgidum* L.) Estimated by SSR, DArT and Pedigree Data. PLoS One 8:e67280. https://doi.org/10.1371/journal.pone.0067280
21. Maccaferri M, Sanguineti · M C, Donini · P, Tuberosa · R (2003) Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. Theor Appl Genet 107:783–797. https://doi.org/10.1007/s00122-003-1319-8
22. Marone D, Laidò G, Gadaleta A, et al (2012) A high-density consensus map of A and B wheat genomes. Theor Appl Genet 125:1619–1638. https://doi.org/10.1007/s00122-012-1939-y
23. Martos V, Royo C, Rharrabti Y, Garcia Del Moral LF (2005) Using AFLPs to determine phylogenetic relationships and genetic erosion in durum wheat cultivars released in Italy and Spain throughout the 20th century. F Crop Res 91:107–116. https://doi.org/10.1016/j.fcr.2004.06.003
24. Marzario S, Logozzo G, David JL, et al (2018) Molecular genotyping (Ssr) and agronomic phenotyping for utilization of durum wheat (*Triticum durum* desf.) ex situ collection from southern italy: A combined approach including pedigreed varieties. Genes (Basel) 9:1–20. https://doi.org/10.3390/genes9100465
25. Moragues M, Moralejo M, Sorrells ME, Royo C (2007) Dispersal of durum wheat [*Triticum turgidum* L. ssp. *turgidum* convar. *durum* (Desf.) MacKey] landraces across the Mediterranean basin assessed by AFLPs and microsatellites. Genet Resour Crop Evol 54:1133–1144. https://doi.org/10.1007/s10722-006-9005-8
26. Motzo R, Giunta F (2007) The effect of breeding on the phenology of Italian durum wheats: From landraces to modern cultivars. Eur J Agron 26:462–470. https://doi.org/10.1016/j.eja.2007.01.007
27. Nei M (1978) ESTIMATION OF AVERAGE HETEROZYGOSITY AND GENETIC DISTANCE FROM A SMALL NUMBER OF INDIVIDUALS. Genetics 89
28. Newton AC, Akar T, Baresel JP, et al (2009) Cereal landraces for sustainable agriculture. Sustain Agric 2:147–186. https://doi.org/10.1007/978-94-007-0394-0_10
29. Oliveira HR, Campana MG, Jones H, et al (2012) Tetraploid wheat landraces in the Mediterranean basin: Taxonomy, evolution and genetic diversity. PLoS One 7:. https://doi.org/10.1371/journal.pone.0037063
30. Paux E, Sourdille P, Mackay I, Feuillet C (2012) Sequence-based marker development in wheat: Advances and applications to breeding. Biotechnol. Adv. 30:1071–1088

31. Peakall R, Smouse PE (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research–an update. Bioinformatics 28:2537–2539. https://doi.org/10.1093/bioinformatics/bts460
32. Percival J (1921) The Wheat Plant: a Monograph
33. Röder MS, Korzun V, Wendehake K, et al (1998) A microsatellite map of wheat. Genetics 149:2007–2023
34. Roussel V, Leisova L, Exbrayat F, et al (2005) SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor Appl Genet 111:162–170. https://doi.org/10.1007/s00122-005-2014-8
35. Ruiz M, Giraldo P, Royo C, et al (2012) Diversity and Genetic Structure of a Collection of Spanish Durum Wheat Landraces. Crop Sci 52:2262–2275. https://doi.org/10.2135/cropsci2012.02.0081
36. Sahri A, Chentoufi L, Arbaoui M, et al (2014) Towards a comprehensive characterization of durum wheat landraces in Moroccan traditional agrosystems: Analysing genetic diversity in the light of geography, farmers' taxonomy and tetraploid wheat domestication history. BMC Evol Biol 14:264. https://doi.org/10.1186/s12862-014-0264-2
37. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454
38. Somers DJ, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114. https://doi.org/10.1007/s00122-004-1740-7
39. Tilman D (1998) The greening of the green revolution. Nature 396:211–212
40. Villa TCC, Maxted N, Scholten M, Ford-Lloyd B (2005) Defining and identifying crop landraces. Plant Genet Resour 3:373–384. https://doi.org/10.1079/pgr200591
41. RStudio Team - RStudio. https://www.rstudio.com/products/team/. Accessed 13 Apr 2021

Figures

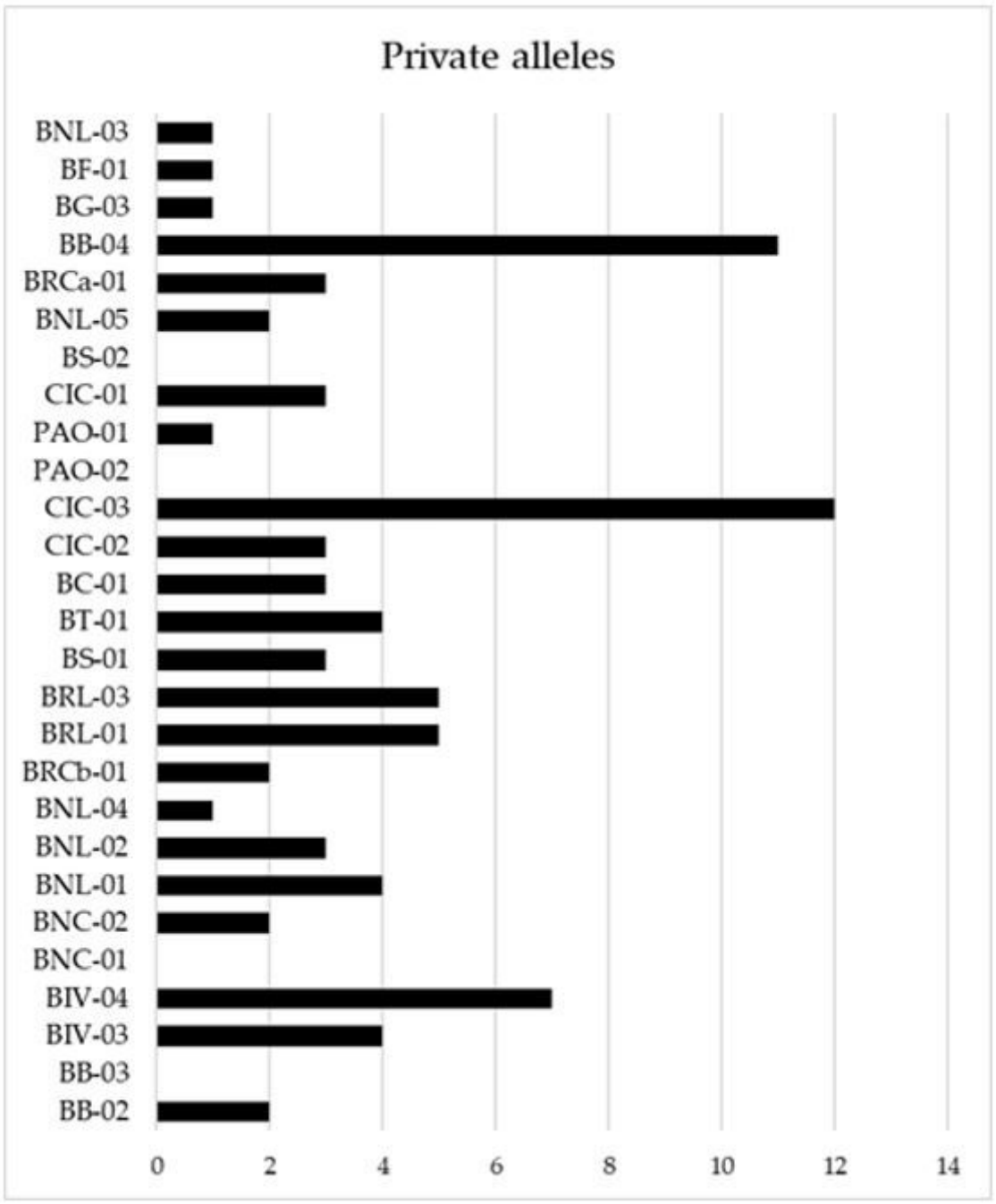

Figure 1

Number of private alleles for each tetraploid wheat

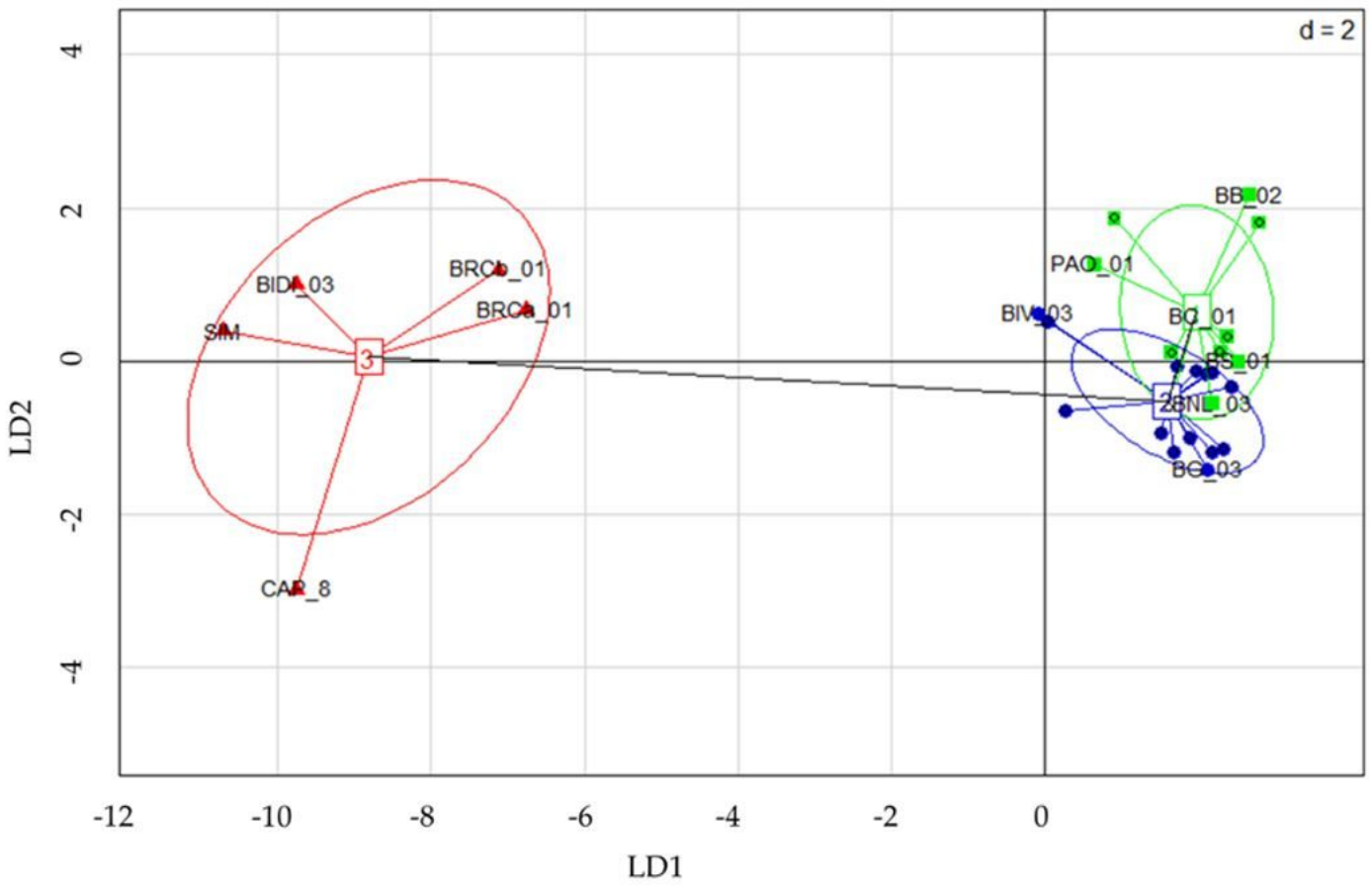

Figure 2

Scatterplot of the first two principal components of the DAPC. Minimum spanning tree connects the three groups. Numbers and colors identify the clusters. LD: loadings

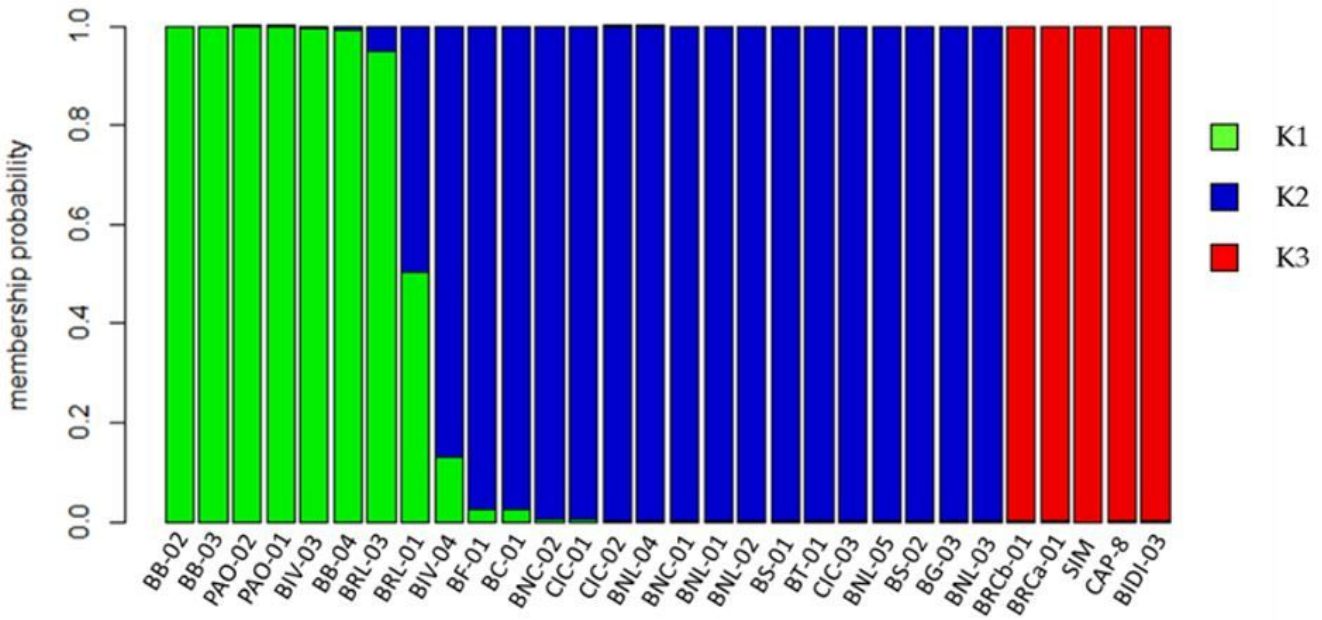

Figure 3

STRUCTURE analysis. Y axis show the membership probability (K-values). Each individual is represented by a vertical line and cluster assignments is indicated by color. Individuals are considered assigned to a cluster if their posterior probability in that cluster is at least 0.7

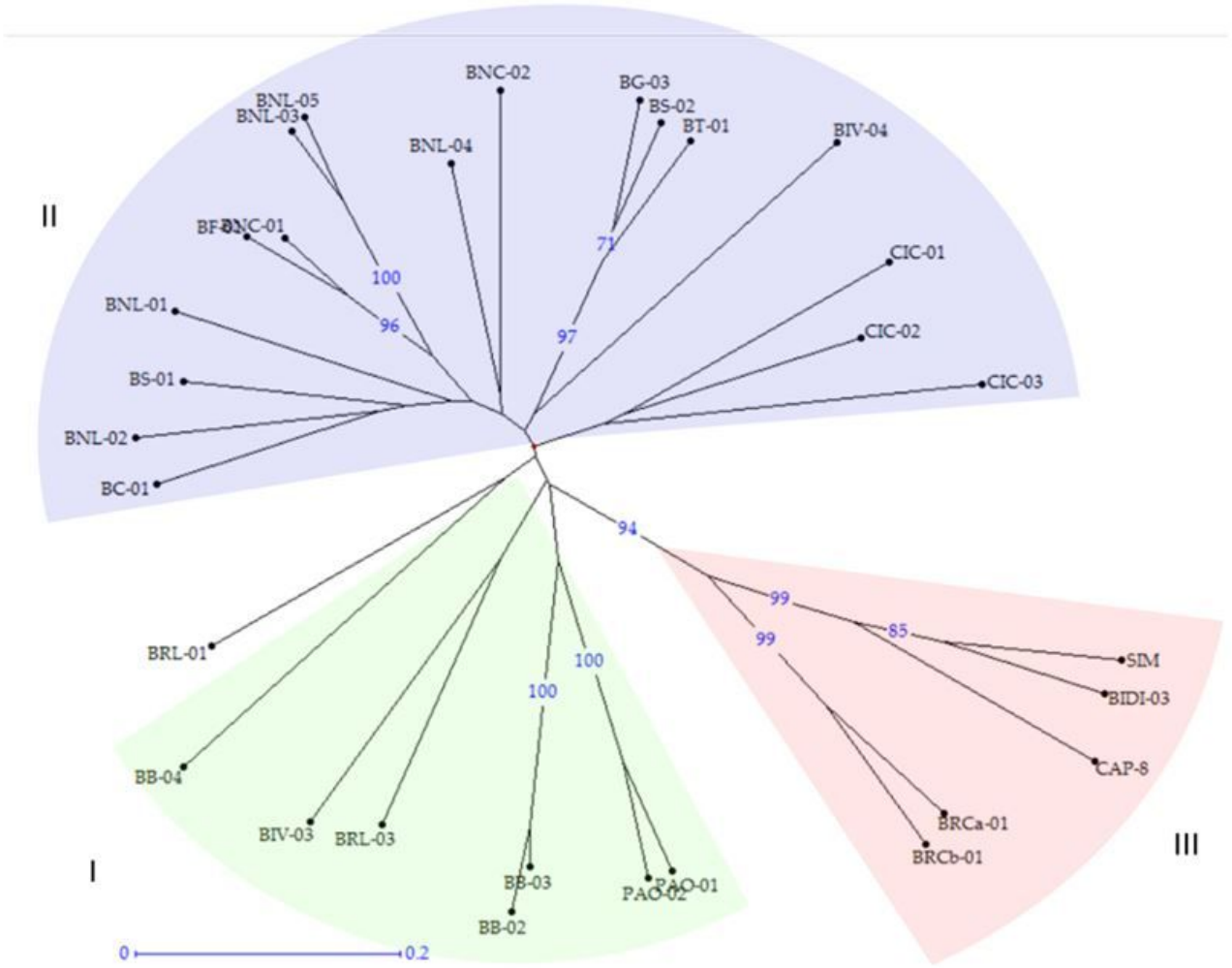

Figure 4

Dendrogram of 30 tetraploid wheats genotypes. Simeto (SIM), Bidì (BIDI-03) and Capeiti (CAP-8) durum wheat varieties were used as outgroup. Dendrogram generated using the neighbor joining method (NJ) and Nei's distance. Groups are indicated by different colors corresponding to STRUCTURE clustering colors and by Roman numbers

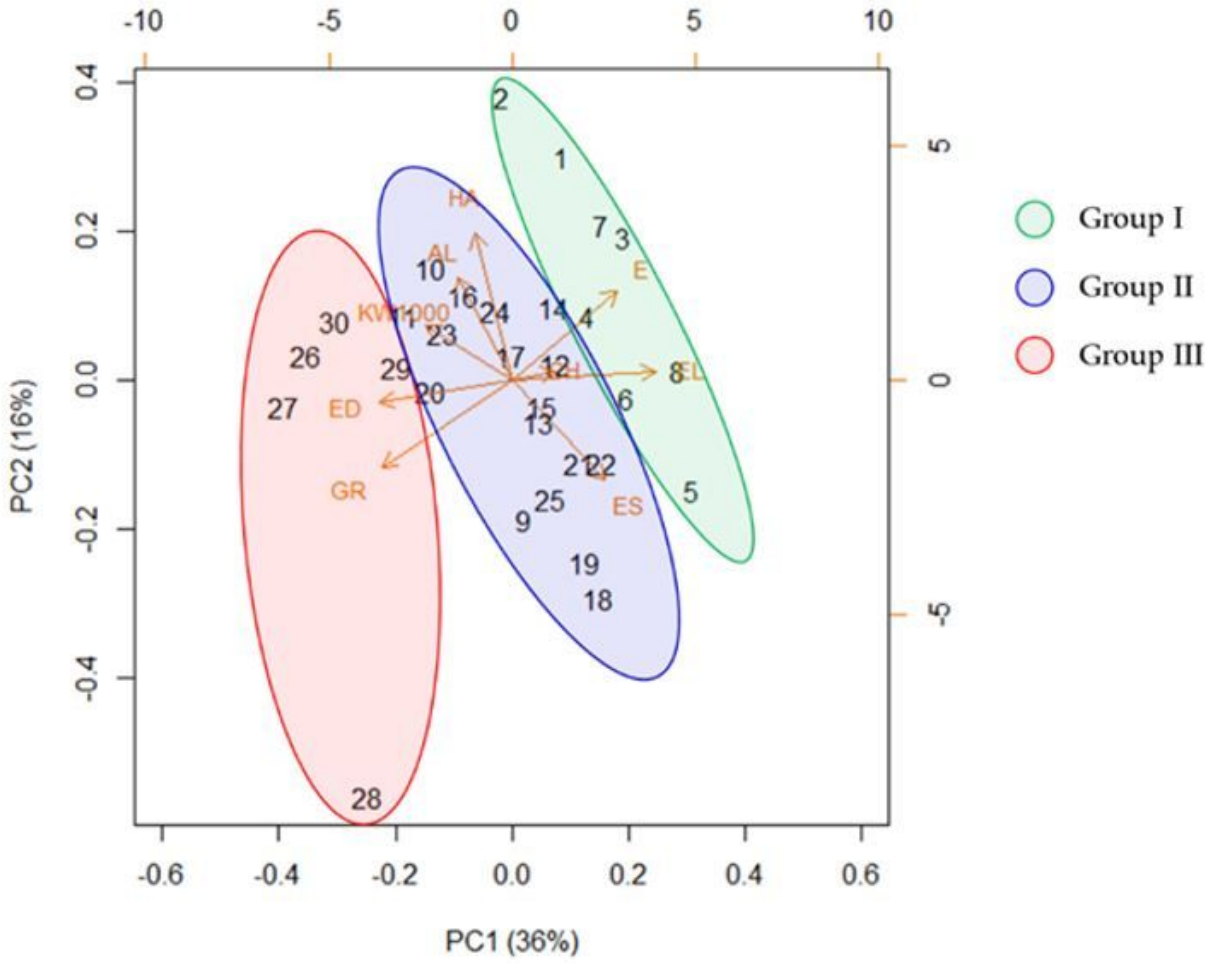

Figure 5

Principal Component Analysis (PCA) of the morpho-agronomic traits in 30 tetraploid wheats. Colored ellipses represent the groups identified in the phylogenetic analysis. Genotypes are numbered from 1 to 30: PAO-02 (1), PAO-01 (2),BB-02 (3), BB-03 (4), BB-04 (5), BRL-03 (6), BIV-03 (7), BRL-01 (8), BIV-04 (9), BNC-01 (10), BNC-02 (11), BNL-01 (12), BNL-02 (13), BNL-04 (14), BC-01 (15), BS-01 (16), BT-01 (17), CIC-02 (18), CIC-03 (19), BS-02 (20), BNL-05 (21), BNL-03 (22), BG-03 (23), BF-01 (24), CIC-01 (25), BIDI-03 (26), SIM (27), CAP-8 (28), BRCb-01 (29), BRCa-01 (30). Traits associated with the samples discrimination are indicated in the plot: Habitus (HA), Time of earing (E), Culm height (CH), Ear length (EL), Awns length (AL), Ear shape (ES), Ear density (ED), 1000-kernel weight (KW1000)

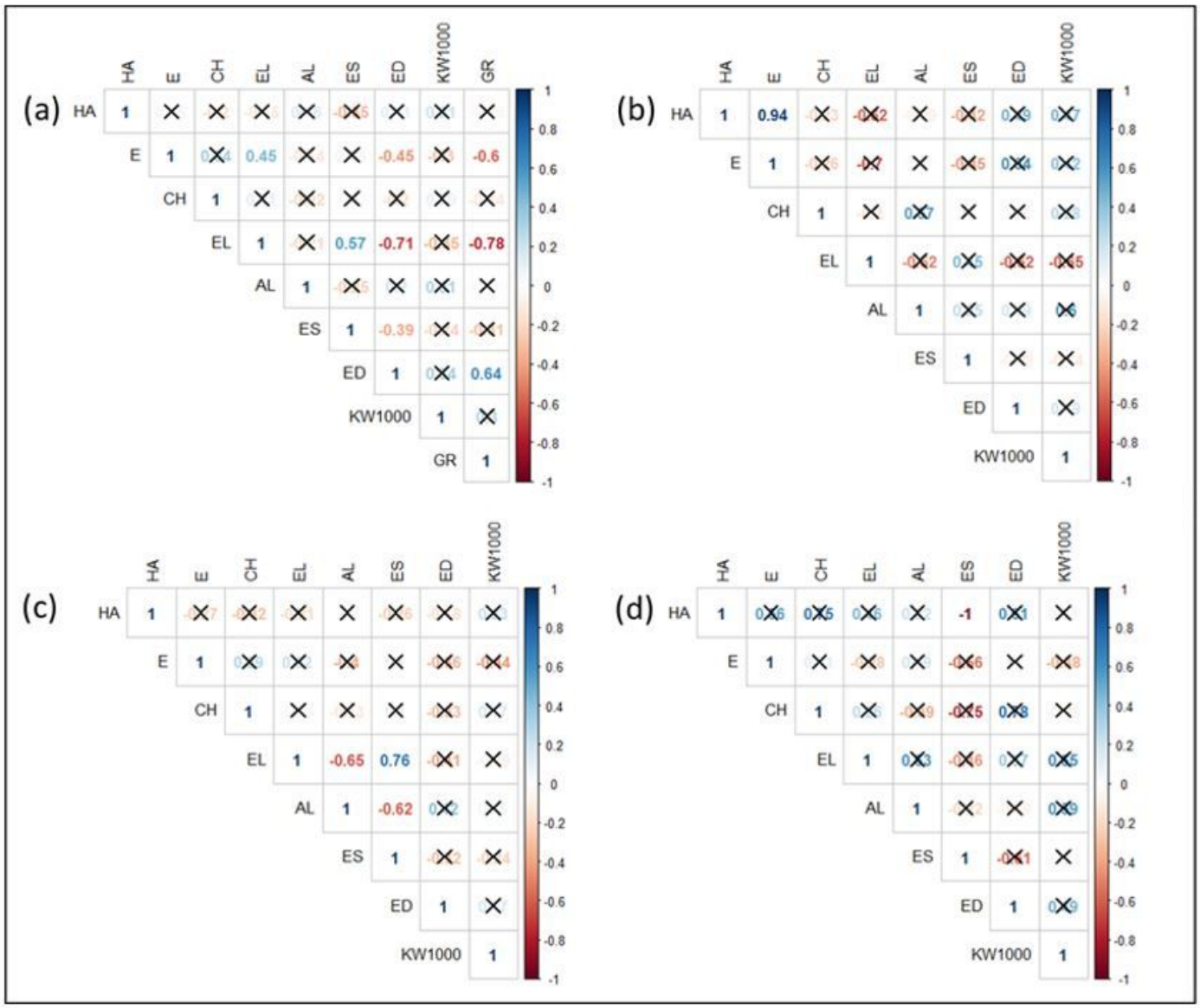

Figure 6

Pearson correlation matrix of 8 morpho-agronomic traits of tetraploid wheats. Numbers indicate the correlation coefficient: positive correlations are displayed in blue and negative correlations in red color. Non-significant correlations ($p>0.05$) are marked by a black cross. (a) traits correlation among all the genotypes (the membership group identified in figure 4 is included in the analysis as “Group” variable); (b) traits correlation among the genotypes of “Group I”; (c) traits correlation among the genotypes of “Group II”; (d) traits correlation among the genotypes of “Group III”. Habitus (HA), Time of earing (E), Culm height (CH), Ear length (EL), Awns length (AL), Ear shape (ES), Ear density (ED), 1000-kernel weight (KW1000)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarymaterial.xlsx