CROP SCIENCE

Diversity and Genetic structure of the Spanish collection of durum wheat (Triticum turgidum L) landraces

Journal:	*Crop Science*
Manuscript ID:	CROP-2012-02-0081-ORA.R2
Manuscript Type:	1. Original Research Articles
Divisions:	C8 plant genetics resources
Date Submitted by the Author:	01-May-2012
Complete List of Authors:	Ruiz, Magdalena; Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Giraldo, Patricia; ETSI Agrónomos, Biotecnología Royo, Conxita; Institute for Food and Agricultural Research and Technology, Field Crops Villegas, Dolors; Institute for Food and Agricultural Research and Technology, Field Crops Aranzana, Mariá; Research Centre (IRTA-CSIC-UAB),, CRAG Agrigenomics Carrillo, Jose; Univ.Politécnica de Madrid (UPM), Biotecnologia
Keywords:	Wheat, Plant genetic resources

Diversity and genetic structure of a collection of Spanish landraces of durum wheat (*Triticum turgidum* L)

Magdalena Ruiz,* Patricia Giraldo, Conxita Royo, Dolors Villegas, M. Jose Aranzana, Jose M. Carrillo

M. Ruiz, Centro Nacional de Recursos Fitogenéticos, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Autovía de Aragón Km 36, 28800 Alcalá de Henares, Spain; P. Giraldo and J.M. Carrillo, Dep. Biotecnología, Escuela Técnica Superior de Ingenieros Agrónomos. Univ. Politécnica de Madrid, Ciudad Universitaria, 28040 Madrid, Spain; C. Royo and D. Villegas, Institute for Food and Agricultural Research and Technology, Field Crops Section, Centre UdL-IRTA, Rovira Roure, 191, 25198 Lleida, Spain; M. J. Aranzana, CRAG Agrigenomics Research Centre (IRTA-CSIC-UAB), Jordi Girona, 18-6, 08034 Barcelona, Spain.
Received_____ *Corresponding author (mruiz@inia.es)

Abbreviations: DArTs, diversity arrays technology markers; F_{ST}, population differentiation; G_{ST}, gene differentiation between populations; H_T, total gene diversity; PIC, polymorphism information content; Pop, population; SSR, short sequence repeat; Z, agro-ecological zones

Keywords DArTs, durum wheat, gliadins, landraces, SSRs, STRUCTURE

ABSTRACT

Knowledge of the genetic structure of germplasm collections is crucial for conservation and efficient use of genetic resources. This study assessed the diversity and genetic structure of a collection of landraces of Spanish durum wheat (*Triticum turgidum* L.) using several marker systems and correlated the diversity and agro-morphological traits with geographic and climatic features. Analyzed genotypes were separated into nine populations with a moderate to great genetic divergence among them. The three subspecies *dicoccon*, *turgidum* and *durum*, present in the collection, largely determined the clustering of the populations. Genotype variation was lower in *dicoccon* and *turgidum* than in *durum*. Genetic differentiation by the agro-ecological zone of origin was greater in *dicoccon* and *turgidum* than in *durum*. DArT markers revealed two geographic substructures, east-west for *dicoccon* and northeast-southwest for *turgidum*. The ssp. *durum* had a more complex structure, consisting of seven populations with high intra-population variation. DArT markers allowed the detection of subgroups within some populations, with agro-morphological and gliadin differences, and distinct agro-ecological zones of origin. Two different phylogenetic groups were detected; revealing that some *durum* accessions were more related to ssp. *turgidum* from northern Spain, while others seem to be more related to *durum* wheats from North Africa.

Introduction

Good understanding of the genetic structure of germplasm collections is critical for conservation *per se*, and for efficient use of genetic resources. A comprehensive assessment of the extent and distribution of genetic variation within endemic populations is essential for their exploitation by breeders and in the development of core collections (van Hintum et al., 2000).

In wheat, Huang et al., (2002); Chao et al., (2007); Li et al., (2008), Royo et al., (2010) and others have demonstrated the effectiveness of the single sequence repeats (SSRs) or microsatellites DNA markers, in studies of germplasm collections using cluster analyses or the model-based clustering implemented in the software STRUCTURE (Pritchard et al., 2000). SSRs offer a number of advantages such as high polymorphism, locus specificity and co-dominance. However, other types of DNA markers such as DArTs (Diversity Arrays Technology) offer a deeper genome coverage and better effectiveness, in the sense that much higher numbers of genotypes can be screened in a time and cost effective manner (Akbari et al., 2006). So, some studies have applied DArT markers to the analyses of genetic diversity and population structure in wheat (Stodart et al., 2007; White et al., 2008; Raman et al., 2010; Zhang et al., 2011; Dreisigacker et al., 2012). Storage proteins such as gliadins can also serve as valuable markers for the quantification of genetic diversity in wheat, and can be used to establish phylogenetic relationships and associations with agro-climatic factors (Kudryavtsev et al., 1996; Melnikova et al., 2010b). All diversity studies have shown that different factors such as selective pressures, pedigree relations, desirable traits for humans, or ecological environments may contribute to the genetic variability of accessions.

Landraces of the Spanish durum wheat (*Triticum turgidum* L.) are classified in three main interfertile subspecies: *dicoccon*, *turgidum* and *durum* that share the same AABB genomic

constitution. The ssp. *dicoccon*, hulled wheat used for animal feed and human consumption, represents the primitive stage in the crop evolution. It was gradually replaced by more advanced free-threshing types of ssp. *durum* and *turgidum*, which evolved from ssp. *dicoccon* (Zohary and Hopf, 1994). In Spain, ssp. *dicoccon* was less widespread, with its cultivation area restricted to mountainous regions. In contrast, the ssp. *durum* was the most widely grown, and was adapted to dry-summer conditions; ssp. *turgidum* was less common than ssp. *durum* and was grown in the areas colder than *durum*. It was a mostly winter wheat with a prostrate growth habit and late heading. Although both subspecies were consumed as pasta and semolina products, Spanish landraces of *turgidum* had, in general, lower quality and higher tillering rate than the *durum* types (Gadea, 1954). Most of these Spanish landraces were collected in the first half of the 20th century, and maintained in the national collection. Several studies have shown that these landraces have a great variability relative to other germplasm collections (Pflüger et al., 2001; Aguiriano et al., 2006, 2008; Moragues et al., 2006). The evaluation of the genetic structure of this collection appears worthwhile for better conservation, utilization, and eventual creation of a core collection, and can be dramatically enhanced by using molecular genotyping tools.

The aim of this research was to study the genetic structure and genetic diversity of a wide sample of Spanish durum wheat landraces selected from the national collection, using SSR, DArT and gliadin-markers. The relationships among the three subspecies taxa present in the collection (*dicoccon*, *turgidum* and *durum*) were assessed and geographic and climatic features, as well as agro-morphological traits, were analyzed in order to obtain information for a more efficient conservation and utilization of the collection.

MATERIALS AND METHODS

Plant Material and Agro-ecological Zones

A total of 190 genotypes of three subspecies: 13 of *T. turgidum* ssp. *dicoccon*, 38 of *T. turgidum* ssp. *turgidum* and 139 of *T. turgidum* ssp. *durum,* were selected for this study from the national collection of 555 Spanish accessions of *T. turgidum* L. maintained at the National Plant Genetic Resources Centre. The *turgidum* accessions were distinguished from *durum* accessions by spike morphology. Spikes of *turgidum* are usually lax and long, with rough awns which can fall off at maturity. They have short glumes and round, soft grain. The *durum* types are always awned, usually smooth at the base, have dense spikes compressed laterally and long grains. The number of accessions selected for each subspecies was proportional to the number of landraces of each subspecies cultivated before 1950 according to Gadea (1954). The set included 185 landraces from all wheat growing areas in Spain as well as five old cultivars. The site from which each accession was collected was ascribed to one of nine agro-ecological zones (Z) in Spain (Fig.1). These zones were defined based on a cluster analysis of historical yield records from 1920 to 1991 in the Spanish provinces (MAPA 1920-1991) following the methodology used by Igartua et al. (1998). Zones with similar yields but different climatic and geographic locations were separated (Z3 and Z4, and Z6 and Z7). The number of landraces representing each zone and province was chosen to be proportional to the number of landraces cultivated in that zone (Gadea, 1954). The selection of accessions within provinces and zones was made using passport data, local names and agro-morphological characterization to represent the diversity in the national collection.

Eighty accessions had geographic coordinates (altitude, latitude and longitude) and climatic data of the collection site. Climatic data from at least 20 years were extracted from the

SIGA service (Spanish acronym for Geographic Information System for Agriculture) of the Spanish Ministry of Agriculture, Food and Environment (http://www.magrama.gob.es/es/cartografia-y-sig/temas/sistema-de-informacion-geografico-de-datos-agrarios-siga-/) (Table 1).

Seed samples were derived from single bagged spikes taken from single selected plants, representative of each original accession. Seeds from bagged spikes were used for gliadin and DArT analysis, and leaves from the same selected plants were used for the SSRs analyses.

Molecular Markers

For the SSR analysis, DNA was extracted from leaf samples using a standard CTAB protocol. The set of 190 accessions was profiled with 39 microsatellites (SSRs) selected based on their genomic distribution, profile quality and the polymorphism level. The list included markers listed as WMC (Wheat Microsatellite Consortium), GWM (Gatersleben Wheat Microsatellite), CFA (Clermont-Ferrand A genome), and BARC (Beltsville Agriculture Research Center) SSRs (Sourdille et al., 2001; Song et al., 2005, Somers et al., 2004) (Table S1). The SSR primer pair sequences and amplification conditions were obtained from the GrainGenes database (http://www.wheat.pw.usda.gov). The forward primer of each marker was 5'-labelled with a fluorescence tag, PCR products were analyzed on ABI PRISM® 3100 DNA analyzer (Applied Biosystems, USA) and allele sizing was carried out by *Peak Scanner Software* v1.0 package (Applied Biosystem, USA).

Gliadins were extracted from single seeds and fractionated in the polyacrylamide gel electrophoresis (Lafiandra and Kasarda, 1985). Identification of *Gli-* alleles was performed

following the nomenclature proposed by Kudryavtsev et al. (1996) and Aguiriano et al. (2006, 2008). The alleles not catalogued in previous studies were termed as 'new'.

For DArT genotyping, DNA was extracted from seeds using the *Ultra Clean Plant DNA Isolation kit* (MO BIO), according to the manufacturer's instructions, and sent for analysis to Diversity Arrays Technology Pty Ltd (http://www.triticarte.com.au/) using the standard Durum Wheat v2.0 array (with about 2,000 markers). From the 190 initial accessions, data were collected for 184 accessions, due to logistical problems and analysis failures in six samples. In total, 749 markers were polymorphic in the collection.

Correlations between genetic distances obtained with molecular markers were tested using the Mantel test (Mantel, 1967).

Analysis of the Genetic Structure of the Collection

The genetic structure of the collection was investigated by SSRs using the Bayesian clustering algorithm implemented in the software STRUCTURE v 2.1 (Pritchard et al., 2000). This model delineates clusters of individuals based on their genotypes at multiple loci by identifying groups of genotypes, termed "populations", with distinctive allele frequencies in which the Hardy–Weinberg equilibrium is maximized. In this study, each genotype corresponds to an accession, so the populations obtained with the analysis represent groups of accessions. The program was run independently five times with k ranging from 1 to 13 in each run using a burn-in of 100,000, run length of 1000,000 and a model allowing for admixture and allele frequencies that are correlated among populations. The five independent runs yielded consistent results. The true number of populations (k) was determined by means of both an estimate of the posterior probability of the data for a given k (as proposed by Pritchard et al., 2000) and the Evanno's Δk (Evanno et al.,

2005) (Fig. 2A, B). A genotype was considered to belong to a population if its membership coefficient was ≥ 0.50 (Royo et al., 2010).

To investigate population differentiation, F_{ST} (Weir and Cockerham, 1984) between populations was calculated and tested using FSTAT 2.9.3.2 (Goudet, 2001). The total gene diversity (H_T) and the relative magnitude of gene differentiation between populations (G_{ST}) were calculated (Nei, 1973) with POPGENE 1.32 (Yeh and Boyle, 1997). For the DArT data analysis, dendrograms were made by UPGMA aggregation method with the Dice distance (Dice, 1945).

Agro-morphological Characterization

The accessions were sown in an augmented design (Petersen, 1985) of four blocks with 54 plots per block during the season 2006-07 in Alcala de Henares (Madrid). Each accession was cultivated in a single plot consisting of 3 rows 1.5 m in length with 30 plants per row. Cultivars Senatore Cappelli, Cocorit 71, Simeto, Don Pedro, Vitron, and Claudio, with different agro-morphological characteristics, were included in each block as checks to estimate the adjustment factor for each block. Agro-morphological traits listed in Tables 2 and 3 were recorded according to IBPGR (1985) from five different plants of each accession. One-hundred and thirty-nine landrace accessions were classified according to their seasonal growth habit (winter, alternative or spring) using the data of Gadea (1954). Differences of trait means among populations were checked by the LSD of the ANOVA for quantitative data and χ^2 test for frequency data. Relationships between variables were examined using Pearson correlation coefficients.

RESULTS

Polymorphism of SSR, gliadin and DArT markers

Among 190 accessions in the study, the 39 SSR markers used identified 641 alleles. The number of alleles per locus ranged from 4 to 41, with the mean of 16 alleles per locus (Table S1). With two exceptions (Xgwm0002 and Xgwm0095), all loci presented unique alleles. Mean value for the polymorphic information content (PIC) was 0.77. The number of accessions with unique alleles was notable, 10 out of 13 (77%) in *dicoccon*, 27 out of 38 (71%) in *turgidum* and 57 out of 139 (41%) in *durum*.

The overall gene diversity for gliadins among the 190 genotypes was 0.80. The most polymorphic loci were *Gli-A2* and *Gli-B2* ($H_T > 0.87$). The three subspecies displayed unique alleles and 23 uncatalogued alleles were detected with low frequencies (Table 4). DArT genotyping identified 749 polymorphic loci, of which 554 were used for subsequent analyses, selected for having less than 5% of missing data. These markers were distributed among all seven homologous groups of wheat chromosomes: with 87, 76, 93, 46, 56, 98, and 78 markers per group 1 to 7, respectively (according to the information placed on: http://www.triticarte.com.au/content/further_development.html). Twenty-eight DArT markers were not assigned to any chromosome. PIC values ranged from 0.03 to 0.50 (the max. value for a biallelic marker), with a mean value of 0.30. The Mantel test detected significant correlations ($P < 0.01$) between the distances obtained with SSRs and gliadins (r = 0.36), SSRs and DArTs (r = 0.50,) and SSRs and agro-morphological data (r = 0.21).

Genetic Structure of the Collection

STRUCTURE gave an optimum population (Pop) number (*k*) of nine (Fig. 2). The assignment of genotypes into populations was consistent among the different runs. All but two genotypes of ssp. *turgidum* and eight of *durum* were grouped into one of the nine populations. All the

populations (i.e. group of accessions) contained one subspecies each, except Pop 4 and Pop 5 (Fig. 2C). The three subspecies clustered separately with DArT markers, with the subspecies *turgidum* being closer to *dicoccon* than to *durum* (Fig. 3).

Genetic differentiation among subspecies, agro-ecological zones and populations quantified with SSRs and gliadins was significant (Table 5). According to Wright's classification (Wright, 1978), pairwise F_{ST} values were high between *dicoccon* and *durum*, but moderate between *dicoccon* and *turgidum* and between *durum* and *turgidum*. For each of the three subspecies, the within-zone variability was greater than the between-zone variability (G_{ST} < 0.43). The differentiation among the nine populations was high and higher than among subspecies and zones (Table 5). The inter-population variability was higher for the subspecies (G_{ST} = 0.79) than for the zone (G_{ST} = 0.17). All population pairwise F_{ST} values were significant (Table 6). According to these values, Pop 1 of *dicoccon,* was the most different population. The three *turgidum* populations, Pops 4, 5 and 7 (Fig. 2C) had moderate differentiation, although gene diversity detected with SSRs and gliadins, respectively, was higher in Pop 4 (H_T = 0.71 and 0.79) than in Pop 7 (H_T = 0.59 and 0.60). For the seven *durum* populations, Pops 2, 3 and 6 had the highest F_{ST} values, whereas Pops 8 and 9 showed the least differentiation. The greatest gene diversity was for Pops 5, 8 and 9 (in the range of 0.62 - 0.72) and the lowest for Pops 2, 3 and 6 (in the range of 0.47-0.60).

Differences between Populations due to Geographic and Climatic Features

All *dicoccon* accessions came from the agro-ecological zones Z9, Z4 and Z1 (Table 7), located in northern Spain (Fig. 1). The values of the climatic factors at the sites of origin showed significant differences between Pop 1 and the remaining populations (Table 1).

Accessions of ssp. *turgidum* in Pop 4 originated from the North and East of Spain (Table 7, Fig. 1). By contrast, those included in Pop 7 originated mainly from the South and West of Spain, and none originated from Z9. Accessions from Z4 (Central Spain) were present in both populations (Pops 4 and 7), but those in Pop 4 originated mostly from the eastern provinces, whereas those in Pop 7 were from the western provinces. Accessions in Pop 4 were collected in the areas with a shorter dry-period and lower mean maximum temperatures of the hottest month than those in Pop 7 (Table 1) in agreement with the north-south distribution of Pops 4 and 7.

Most accessions of ssp. *durum* originated from areas of Central or Southern Spain, with a high frequency of specific agro-ecological zones (Table 7, Fig. 1). Few differences were found for the climatic and geographic factors between the major populations, Pops 2, 3, 6, 8, and 9 (Table 1). By contrast, Pop 4 came from more humid and colder areas, and Pop 5 came mostly from colder areas. Both these populations differed from the other durum populations in the absence or higher presence of some agro-ecological zones, indicating that Pops 4 and 5 were spread across different zones than others (Table 7).

Differences between Populations in Agro-morphological Data, Gliadins and DArT Markers

The agro-morphological data indicated that all Spanish *dicoccon* accessions had rough awns, hairless glumes, and red grains (Table 2). The genotype in Pop 5 had dense spikes, the lowest grain weight and a different gliadin pattern from those genotypes in Pop 1 (Tables 2, 3 and 4). Fig. 4 shows the dendrogram based on DArT markers for the *dicoccon* genotypes. The *dicoccon* of Pop 5 was separated from those included in Pop 1. The genotypes of Pop 1 formed three subpopulations, A1, A2 and A3, with corresponding differences in agro-ecological zones of origin, different *Gli-1* alleles and the agro-morphological traits such as growth habit, awnedness,

precocity, and 100-grain weight. There were only two *dicoccon* accessions with spring growth habit (from Z4), in contrast to the winter or alternative growth class of the remaining accessions.

For ssp. *turgidum*, significant ($P < 0.05$) differences between Pops 4 and 7 were detected for *Gli-A1* alleles (Table 4), growth habit and seed color (Table 2), and for quantitative data (Table 3). For Pop 4, 86% and 5% of the accessions were winter and spring types, respectively. By contrast, 33% and 42% of the accessions in Pop 7 were winter and spring types, respectively. Significant correlations were observed between the longitude of the collection site and the number of spikelets per spike ($r = 0.62$, $P < 0.01$), and between latitude and days to heading ($r = 0.58$, $P < 0.05$), days to maturity ($r = 0.63$, $P < 0.01$) and spike length ($r = -0.49$, $P < 0.05$). Earliness and long spikes were associated with southern collection sites, while greater spikelet numbers were associated with the western origin. Cluster analysis with DArT data showed that the genotypes of Pop 4 were separated from those of Pop 7 (Fig. 5). Three subdivisions appeared in Pop 4 according to the agro-ecological zones.

For ssp. *durum*, significant differences ($P < 0.01$) among populations were detected at all gliadin loci (Table 4), for all qualitative traits except for awnedness (Table 2), and for quantitative traits (Table 3). No relationships were detected between populations and spring or winter growth habits, perhaps because of a high proportion (95%) of spring types in this subspecies. The altitude of the collection site correlated with the number of days to maturity ($r = 0.29$, $P < 0.05$). In the dendrogram obtained from DArT data, the genotypes were separated into seven subgroups (Fig. 6). The analysis of these subgroups within each population indicated that over 80% of the genotypes of Pops 2 and 6 were included in a single subgroup. All genotypes of Pop 3 and about 85% of the genotypes of Pops 4 and 9 were clustered in two subgroups. By contrast, Pops 8 and 5 were more subdivided. Differences among subgroups within Pops 3, 4, 8,

and 9 were found for gliadins, some agro-morphological traits as for glume and seed color, and for agro-ecological zones.

DISCUSSION

The suitability of SSR markers for evaluating genetic relationships of durum wheat germplasm (Eujayl et al., 2002; Maccaferri et al., 2003; Royo et al., 2010) and the power of DArT markers for assessing the population structure and genetic diversity in wheat collections have been demonstrated previously (Stodart et al., 2007; White et al., 2008; Raman et al., 2010; Zhang et al., 2011; Dreisigacker et al., 2012). To our knowledge, no diversity study has been conducted in wheat with both types of markers. In this study, there was a large correspondence between the results obtained with SSRs and with DArTs, reflected by significant correlations observed for the genetic distances obtained with the two systems. These correlations can be found for related cultivars or when linkage disequilibrium exists between the different markers' loci, which could be due to the presence of genes for adaptation to local conditions. Although DArT markers revealed less polymorphism information per locus than SSRs, they are useful in diversity analyses because high numbers of markers can be processed quickly, providing a genome-wide coverage and the power to detect even very small polymorphisms. In this study, the DArT markers were very useful to explore the substructure of the collection and to discriminate between subspecies, as has been recently proved in *Aegilops tauschii* (Sohail et al., 2012).

The level of diversity of 0.77 assessed with SSRs in our collection was greater than that observed in other studies of durum wheat, using some coincident SSRs markers. Collections from different countries such as Italy, Oman, Syria and Ethiopia showed average gene diversity values between 0.55 and 0.68 (Eujayl et al., 2002, Macaferri et al., 2003, Teklu et al., 2006, Al-Khanjari et al., 2007, Achtar et al., 2010). The diversity value of our collection assessed with

gliadins was 0.80, much higher than that found in other studies of durum wheat (0.53 in Kudryavtsev et al., 1996, and 0.55 in Melnikova et al., 2010b) Thus, we conclude that we have studied a sample of *T. turgidum* L. with considerable polymorphism, containing many unique alleles, in which the assessment of population structure was essential.

According to STRUCTURE software, the collection was organized in nine populations or groups of accessions with a moderate to great genetic divergence among them. The subspecies taxa, *dicoccon*, *turgidum* and *durum*, had more effect on the differentiation of the populations than the zones in which the accessions were collected. In fact, seven of the nine populations were composed of only one subspecies (Fig. 2C). This separation among subspecies was consistent when assessed with the gliadin and DArT markers (Table 4, Fig. 3). Genotype differentiation was lower in *dicoccon* (one major population) and *turgidum* (two major populations) than in *durum* (five major populations). The genotypes of *turgidum* showed an intermediate position between those of *durum* and *dicoccon*, more related to *durum* with SSRs (Table 5) and to *dicoccon* with DArTs (Fig. 3). The three subspecies originated from climatically different growing areas with different requirements, especially the ssp. *dicoccon* which came from cool, humid areas of northern Spain, in contrast to *durum*, which came from warmer areas in the South (Tables 1 and 7, Fig. 1). The two populations of *turgidum* came from contrasting environments: Pop 4 came from northern areas, whereas Pop 7 came from warmer southern areas, not unlike those of *durum* (Table 1). Some *durum* genotypes of Pop 4 showed common characteristics with *turgidum* and *dicoccon*, such as their northern geographic origin, prostrate growth habit, lax and long spikes, red seeds, and the presence of the allele *Gli-B1new-1* (Tables 2, 3, 4, and 7). Furthermore, some of them were the *durum* genotypes most closely related to *turgidum* in the DArTs. By contrast, they were more similar to *durum* populations in awn length, grain filling

period and DArTs (Tables 1 and 3, Fig. 6). In fact these genotypes in Pop 4 had characteristics of both subspecies. These results are in agreement with Mac Key (1966), who also found mixtures of *durum* and *turgidum* in primitive endemic farms. For the remaining populations, the three subspecies were clearly separated, based on SSR, gliadin and DArT markers, in agreement with Mac Key (2005), who considered these taxa as subspecies. Moreover, genetic diversity in *turgidum* was affected more by the geographic origin of the accessions and seasonal growth habit than in *durum*. These differences between the two subspecies, attributed to their different growing environments and human selection, were also detected by SSRs, which are under less selection pressure than agronomic characteristics, even though some marker loci can suffer frequency shifts if they are sufficiently closely linked to the genes targeted by selection.

Genetic differentiation by zones was greater in *dicoccon* and *turgidum* than in *durum*. DArT analyses detected an eastern-western substructure in *dicoccon* (Fig. 1 and 4) associated with agro-morphological and gliadin differentiation. In ssp. *turgidum*, differences between the two main populations in the agro-morphological traits, in gliadins, as well as in geographic and climatic features, and DArTs followed a northeast-southwest pattern (Tables 1, 2, 3, 4 and 7, Fig. 5). This geographic pattern was confirmed by the relationships between some agro-morphological traits and the longitude and latitude of the collection sites. Similarly to ssp. *turgidum*, a relationship among precocity, spring types and southern distribution has also been observed among Spanish barley genotypes (Lasa et al., 2001; Yahiaoui et al., 2008). This north-south pattern in barley was related to climatic differences in temperature and humidity (Yahiaoui et al., 2008). The association of the distribution of genetic diversity with geographical patterns in *dicoccon* and *turgidum* highlights the effect of selection pressure favoring allele associations

with a better local adaptation. These differences should be considered in the selection of accessions for breeding programs.

The ssp. *durum* showed a more complex genetic structure. Seven populations or groups of accessions were defined by the software STRUCTURE; high F_{ST} values, differences in agro-morphological traits and gliadins indicated that there was also a significant separation of *durum* populations overlaying a high intra-population variation (Tables 2, 3, 4, and 6). These differences cannot be explained only by climatic and geographic variation (Tables 1 and 7). However, minor climatic differences among the sites of origin of some populations could be the result of a high frequency of some collecting zones in those populations e.g. the higher altitude of the collecting zones of Pop 2 were due to the higher presence of the mountainous Z3 in this population. In addition, Pop 9 was associated with warm and dry areas because included a high number of accessions from Z6, which is the driest Spanish zone (Tables 1 and 7). Furthermore, some populations showed an eastern (Pops 3 and 9) -western (Pop 8) differentiation associated with the presence or absence of Z6 or Z8 (eastern), or Z5 (western) (Table 7, Fig. 1). In spite of these associations with geographic origin, the studied populations also came from different agro-ecological zones, suggesting that genetic diversity was not completely related to the geographic distribution. In addition, genetic variation was larger within than between zones, indicating that genetic diversity was preserved independently of the geographic region of origin of the accessions. A similar result was obtained by Li et al. (2008) and Melnikova et al. (2010a) working with durum wheat landraces from China and Bulgaria, respectively. Different human practices or germplasm exchange could have contributed to the increased variation within populations.

The DArT markers based analysis detected the presence of subgroups within some *durum* populations defined by STRUCTURE. This substructure reflected differences in agro-morphological traits, mainly in spike characteristics, gliadins and the agro-ecological zones of

origin. Populations 2, 6 and 3 were less subdivided with DArTs, suggesting a higher isolation, in agreement with their lower variability and higher differentiation (Table 6, Fig. 6). The accessions in Pop 2 were distinguished by their late-heading and high altitude of collection sites, while those in Pop 3 were characterized by their earliness (Tables 1 and 3). Interestingly, Pop 6 included a distinct agrotype with erect growth habit and smooth awns (Table 2) coming from Z8, mainly from the Balearic Islands. By contrast, Pops 4, 5, 8, and 9 showed more subdivisions, greater variability and less differentiation from other populations, including Pop 1 of *dicoccon* (Table 6, Fig. 6). These results indicate that Pops 4, 5, 8, and 9 are more heterogeneous with higher levels of gene flow, with the subgroup G of Pop 4 being the most closely related to the subspecies *turgidum*.

The lower diversity of *Gli-1*, together with the fact that the alleles with differences among populations had high frequencies, suggests that selection pressure favors alleles associated with good local adaptation. Moreover, *Gli-A1* and *Gli-B1* were the most fixed loci within zones in ssp. *turgidum* and *dicoccon* (G_{ST} = 0.45 and 0.59, respectively) in agreement with other studies that have reported that some *Gli-1* alleles may provide some advantage under specific agro-climatic conditions (Melnikova et al., 2010a,b). The lower diversity of *Gli-B1* could also be due to a selection for quality, since these gliadins are genetic markers for gluten quality (Damidaux et al., 1978). The *Gli-B1new-1* allele, widespread in Pop 4, was also the most common in *turgidum* and very frequent in *dicoccon* (Table 4), confirming previous observations in accessions from northern Spain (Aguiriano et al., 2008). This allele, associated with poorer quality than alleles *b* and *c* (Aguiriano et al., 2009), has been reported to be very specific to Spanish germplasm (Kudryavtsev et al., 1996; Melnikova et al., 2010b). In the *Gli-A2* locus, the allele *o*, present in the old cultivar of Algerian origin, Senatore Cappelli, was very frequent in subgroup B of Pop 3,

to which this cultivar belongs. This allele was absent in ssp. *dicoccon*, ssp. *turgidum* and in Pop 4 of ssp. *durum*, i.e. in landraces from the northern areas (Tables 4 and 7, Fig. 1), supporting the hypothesis of a North-African origin of this allele (Kudryavtsev et al., 1996). Similarly, *Gli-B2h* of Senatore Cappelli, very common in Pop 2, subgroup B of Pop 3 and Pop 6 (Table 4), was more frequent in the south and east of the country, in the Canary Islands, near North Africa, and absent in the North of Spain. High frequencies of *Gli-A2o* were observed in cultivars of Algeria, Morocco and Tunisia, and of *Gli-B2h* in many countries of Africa and South West Asia, with the highest occurrence in cultivars from Algeria (Melnikova et al., 2010b).

A relationship between the SSR classification and the pedigree of wheat lines has been shown in some studies (Macaferri et al., 2003; Chao et al., 2007; Royo et al., 2010). Our results seem to indicate that genotypes in the subgroup B of Pop 3 and some genotypes of Pops 2 and 6 could share related ancestral lines from North Africa. Moragues et al. (2007) also found that landraces from the Iberian Peninsula and North Africa (Egypt, Algeria, and Morocco) group together, based on AFLPs and SSRs. Our results support the hypothesis of Moragues et al. (2007) of two dispersal patterns of durum wheat across the Mediterranean basin, one along the north side (Syria, Turkey, Bulgaria, Cyprus, Greece, and Italy) and the second along the south (Egypt, Algeria, Morocco, Spain, and Portugal).

The genetic structure for the Spanish germplasm of *T. turgidum* L. showed that the subspecies was the first level of separation, followed by the agro-ecological zone of origin for ssp. *dicoccon*, but by geographic and seasonal growth habit for ssp. *turgidum*. Landraces of ssp. *durum* showed a complex variation pattern, including different phylogenetic groups. Some accessions were more related to the ssp. *turgidum* from northern Spain, while others, markedly different, less diverse and more isolated, seem to be more related to *durum* wheats from North

Africa. Natural and human selection in combination with human migrations may have contributed to the genetic diversity of *durum* wheat Spanish landraces. The knowledge of the genetic and geographical structure of the collection is valuable for selecting accessions by users and essential for the creation of the core collection by hierarchical sampling methods. The detection of redundancies and missing traits, such as the absence of hairy and black spikes in *dicoccon*, or the convenience of separation of ssp. *durum* and *turgidum* in evaluation studies help to improve the conservation of the collection.

ACKNOWLEDGMENTS

This research was supported by the project RF2006-00020-C03-01 from the National Institute for Agricultural and Food Research and Technology of Spain, the projects AGL-2009-11187 and AGL-2009-9980 from the Ministry of Science and Innovation, and the European Fund for Regional Development (FEDER). We thank Rosario Fité and Mª Jesús Tomás for their technical assistance. The Centre University of Lleida- Institute for Food and Agricultural Research and Technology forms part of the Centre CONSOLIDER INGENIO 2010 on Agrigenomics funded by the Spanish Ministry of Education and Science.

REFERENCES

Achtar, S., M.Y. Moualla, A. Kalhout, M.S. Roder, and N. MirAli. 2010. Assessment of genetic diversity among Syrian durum (*Triticum* ssp. *durum*) and bread wheat (*Triticum aestivum* L.) using SSR markers. Russ. J. Genet. 46:1320-1326.

Aguiriano, E., M. Ruiz, R. Fite, and J.M. Carrillo. 2006. Analysis of genetic variability in a sample of the durum wheat (*Triticum durum* Desf.) Spanish collection based on gliadin markers. Genet. Resour. Crop Evol. 53:1543-1552.

Aguiriano, E., M. Ruiz, M.R. Fité, and J.M. Carrillo. 2008. Genetic variation for glutenin and gliadins associated with quality in durum wheat (*Triticum turgidum* L. ssp. *turgidum*) landraces from Spain. Span. J. Agric. Res. 6:599-609.

Aguiriano, E., M. Ruiz, R. Fité, and J.M. Carrillo. 2009. Effects of N fertilisation, year and prolamin alleles on gluten quality in durum wheat (*Triticum turgidum* L. ssp. *turgidum*) landraces from Spain. Span. J. Agric. Res. 7:342-348.

Akbari, M., P. Wenzl, V. Caig, J. Carling, L. Xia, S. Yang, et al. 2006. Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor. Appl. Genet. 113:1409-1420.

Al-Khanjari, S., K. Hammer, A. Buerkert, and M.S. Roder. 2007. Molecular diversity of Omani wheat revealed by microsatellites: I. Tetraploid landraces. Genet. Resour. Crop Evol. 54:1291-1300.

Chao, S., W. Zhang, J. Dubcovsky, and M. Sorrells. 2007. Evaluation of genetic diversity and genome-wide linkage disequilibrium among US wheat (*Triticum aestivum* L.) germplasm representing different market classes. Crop Sci. 47:1018-1030.

Damidaux, R., J.C. Autran, P. Grignac, and P. Feillet. 1978. Mise en évidence de relations applicables en sélection entre l'électrophorégramme des gliadines et les propriétés viscoélastiques du gluten de *Triticum durum* Desf. CR Acad.Sc.Paris, Série D 287:701-704.

Dreisigacker, S., H. Shewayrga, C. Crossa, V.N. Arief, I.H. DeLacy, R.P. Singh, et al. 2012. Genetic structures of the CIMMYT international yield trial targeted to irrigated environments. Mol. Breeding 29:529-541.

Dice, L R. 1945. Measures of the amount of ecologic association between species. Ecology 26:297–302.

Eujayl, I., M.E. Sorrells, M. Baum, P. Wolters, and W. Powell. 2002. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. Theor. Appl. Genet. 104:399-407.

Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. Mol. Ecol. 14:2611-2620.

Gadea, M. 1954. Trigos españoles. Instituto Nacional de Investigaciones Agronómicas, Madrid, España.

Goudet, J. 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (Version 2.9.3). Available from http://www2.unil.ch/popgen/softwares/fstat.htm

Huang, X., A. Börner, M. Röder, and M. Ganal. 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor. Appl. Genet. 105:699-707.

IBPGR. 1985. Revised Descriptor list for wheat (*Triticum* ssp). International Board for Plant Genetic Resources, Rome, Italy.

Igartua, E., M.P. Gracia, J.M. Lasa, B. Medina, J.L. Molina-Cano, J.L. Montoya, et al. 1998. The Spanish barley core collection. Genet. Resour. Crop Evol. 45:475-481.

Kudryavtsev, A.M., G. Boggini, S. Benedettelli, N.N. Illichevskii. 1996. Gliadin polymorphism and genetic diversity of modern Italian durum wheat. J. Genet. Breed. 50:239-248

Lafiandra, D., and D.D. Kasarda. 1985. One-and two-dimensional (two-pH) polyacrylamide gel electrophoresis in a single gel: Separation of wheat proteins. Cereal Chem. 62:314-319.

Lasa, J.M., E. Igartua, F.J. Ciudad, P. Codesal, E.V. Garcia, M.P. Gracia, et al. 2001. Morphological and agronomical diversity patterns in the Spanish barley core collection. Hereditas 135:217–225.

Li, W., P. Dong, Y.M. Wei, G.Y. Cheng, and Y.L. Zheng. 2008. Genetic variation in *Triticum turgidum* L. ssp. *turgidum* landraces from China assessed by EST-SSR markers. Agricultural Sciences in China 7:1029-1036.

Mac Key, J. 1966. Species relationship in *Triticum*. Hereditas 2:237-276.

Mac Key, J. 2005. Wheat: Its concept, evolution and taxonomy. Durum Wheat Breeding. In: Royo C, et al., editors, Durum wheat breeding: current approaches and future strategies. Haworth Press, New York. vol 1, p 3-61

Maccaferri, M., M.C. Sanguineti, P. Donini, and R. Tuberosa. 2003. Microsatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. Theor. Appl. Genet. 107:783-797.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209.

Ministerio de Agricultura, Pesca y Alimentación (MAPA). 1920-1991. Anuario de Estadística Agraria. Subdirección General de Estadística, Servicio de Estadísticas Agrarias, Madrid.

Melnikova, N.V., G.D. Ganeva, Z.G. Popova, S.P. Landjeva, and A.M. Kudryavtsev. 2010a. Gliadins of Bulgarian durum wheat (*Triticum durum* Desf.) landraces: genetic diversity and geographical distribution. Genet. Resour. Crop Evol. 57:587-595.

Melnikova, N.V., O.P. Mitrofanova, O.A. Liapounova, and A.M. Kudryavtsev. 2010b. Global diversity of durum wheat *Triticum durum* Desf. for alleles of gliadin-coding loci. Russian J. of Genet. 46:43-49.

Moragues, M., M. Moralejo, M.E. Sorrells, and C. Royo. 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.

Moragues, M., J. Zarco-Hernández, M.A. Moralejo, and C. Royo. 2006. Genetic diversity of glutenin protein subunits composition in durum wheat landraces [*Triticum turgidum* ssp. *turgidum* convar. *durum* (Desf.) MacKey] from the Mediterranean basin. Genet. Resour. Crop Evol. 53:993-1002.

Nei, M. 1973. Analysis of gene diversity in subdivided populations. Proceedings of the National Academy of Sciences 70:3321.

Petersen, R.G. 1985. Augmented designs for preliminary yield trials (revised). Rachis 4:27-32.

Pflüger, L.A., L.M. Martin, and J.B. Alvarez. 2001. Variation in the HMW and LMW glutenin subunits from Spanish accessions of emmer wheat (*Triticum turgidum* ssp. *dicoccum* schrank). Theor. Appl. Genet. 102:767-772.

Pritchard, J.K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945.

Raman, H., B. J. Stodart, C. Cavanagh, M. Mackay , M. Morell, A. Milgate, et al. 2010. Molecular diversity and genetic structure of modern and traditional landrace cultivars of wheat (*Triticum aestivum* L.). Crop Pasture Sci. 61:222-229.

Royo, C., M. Maccaferri, F. Álvaro, M. Moragues, M.C. Sanguineti, R. Tuberosa, et al. 2010. Understanding the relationships between genetic and phenotypic structures of a collection of elite durum wheat accessions. Field Crops Res. 119:91-105.

Sohail Q., T. Shehzad, A. Kilian, A.E. Eltayeb, H. Tanaka, and H. Tsujimoto. 2012. Development of diversity array technology (DArT) markers for assessment of population structure and diversity in *Aegilops tauschii*. Breed. Sci. 62:38-45.

Somers, D.J., P. Isaac, and K. Edwards. 2004. A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 109:1105-1114.

Song, Q.J., J.R. Shi, S. Singh, E.W. Fickus, J.M. Costa, J. Lewis, et al. 2005. Development and mapping of microsatellite (SSR) markers in wheat. Theor. Appl. Genet. 110:550-560.

Sourdille, P., M. Tavaud, G. Charmet, and M. Bernard. 2001. Transferability of wheat microsatellites to diploid Triticeae species carrying the A, B and D genomes. Theor. Appl. Genet. 103:346-352.

Stodart, B.J., M.C. Mackay, and H. Raman. 2007. Assessment of molecular diversity in landraces of bread wheat (*Triticum aestivum* L.) held in an ex-situ collection with Diversity array Technology (DArT ™). Austr. J. Agron. Res. 58:1174-1182.

Teklu, Y., K. Hammer, X.Q. Huang, and M.S. Roder. 2006. Analysis of microsatellite diversity in Ethiopian tetraploids wheat landraces. Genet. Resour. Crop Evol. 53:1115-1126.

Van Hintum, T.J.L., A.H.D. Brown, C. Spillane, and T. Hodgkin. 2000. Core collections of plant genetic resources. IPGRI Technical Bulletin No.3. International Plant Genetic Resources Institute, Rome, Italy

Weir, B.S., and C.C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38:1358-1370.

White J., J.R. Law, I. Mackay, K.J. Chalmers, J.S.C. Smith, A. Kilian, et al. 2008. The genetic diversity of UK, US and Australian cultivars of *Triticum aestivum* measured by DArT markers and considered by genome. Theor. Appl. Genet. 116:439-453.

Wright, S. 1978. Evolution and the genetics of population. Vol 4. Variability within and among natural populations. . University of Chicago Press, Chicago.

Yahiaoui, S., E. Igartua, M. Moralejo, L. Ramsay, J.L. Molina-Cano, F.J. Ciudad, et al. 2008. Patterns of genetic and eco-geographical diversity in Spanish barleys. Theor. Appl. Genet. 116:271-282.

Yeh, F.C., and T.J.B. Boyle. 1997. Population genetic analysis of codominant and dominant markers and quantitative traits. Belg. J. Bot. 129:157.

Zhang, L.Y., D.C. Liu, X.L. Guo, W.L. Yang, J.Z. Sun, D.W. Wang, et al. 2011. Investigation of genetic diversity and population structure of common wheat cultivars in northern China using DArT markers. BMC Genetics 12:42.

Zohary, D., and M. Hopf. 1994. Domestication of Plants in the Old World. Oxford University Press, New York.

FIGURE CAPTIONS

Figure 1. Map of Spanish provinces showing the nine agro-ecological zones for durum wheat cultivation according to long-term (1920-1991) average yield at each province.

Figure 2. (A) Changes in the natural log probability of the data against the number of populations (k). (B) Average Δk against k. Values are the mean of 5 runs of STRUCTURE simulations. (C) Inferred structure of the durum wheat collection based on 190 genotypes and genomic SSR markers using STRUCTURE. Each individual is represented by a line divided into colored segments that represent the individual's estimated membership fractions to each of the nine populations. The subspecies are indicated as *dic* = *dicoccon*, *dur* = *durum* and *tur* = *turgidum*.

Figure 3. Dendrogram of the DArT genotypes of the 184 accessions analyzed.

Figure 4. Dendrogram of the DArT genotypes of the ssp. *dicoccon.* The populations (Pop), DArT subgroup (A) and agro-ecological zone (Z) are indicated.

Figure 5. Dendrogram of the DArT genotypes of the spp. *turgidum.* The populations (Pop) and agro-ecological zone (Z) are indicated.

Figure 6. Dendrogram of the DArT genotypes of the ssp. *durum*. The DArT subgroups (A-G) are outlined in bold. The table shows the percentage of each population (Pop) classified in a DArT subgroup.

Table 1 Means and standard deviations (SD) of altitude and climatic factors of the collecting sites of the accessions contained in the nine populations (Pop).

	Altitude (m)		ETP (mm)		Hot period (months) ‡		Cold period (months) ∫		Dry period (months) ¶		Annual rainfall (mm)		Annual temperature (°C)		Max. temperature of the hottest month (°C)		Min. temperature of the coldest month (°C)	
Pop	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	884.43 a	363.20	627.14 c	48.03	0.00 a	0.00	7.89 a	1.23	0.74 a	0.62	1056.81 a	69.43	10.23 a	1.79	22.83 a	0.81	0.76 a	1.54
2	683.58 ab	327.97	840.00 b	77.46	2.58 cd	1.08	4.93 bc	1.81	4.31 c	1.43	564.32 bc	176.07	15.80 c	1.69	34.33 d	1.53	2.24 ab	2.23
3	440.89 bc	283.46	848.11 b	68.65	2.78 d	1.09	5.00 bc	1.41	4.17 c	0.43	521.71 bc	84.36	15.92 c	1.65	34.37 d	1.72	2.31 ab	1.84
4	576.43 bc	299.78	731.50 a	91.33	0.71 ab	1.20	5.99 b	1.35	2.61 b	1.60	704.01 b	338.64	13.19 b	2.45	28.56 b	3.89	1.39 ab	2.41
5	417.75 bc	395.56	796.75 ab	109.35	1.50 bc	1.29	4.70 bc	1.47	3.15 bc	1.49	731.20 bc	352.53	15.10 bc	2.69	30.68 bc	5.40	3.55 b	1.88
6	629.67 abc	66.12	771.00 ab	119.20	2.00 bcd	1.73	5.97 abc	0.95	3.10 bc	1.77	510.53 bc	162.73	14.17 bc	2.97	30.80 bcd	4.42	0.17 a	1.89
7	520.80 bc	183.89	785.20 ab	45.53	1.60 bcd	0.89	5.22 bc	1.07	4.19 c	0.47	475.18 bc	106.02	14.58 bc	1.34	32.50 cd	1.89	1.96 ab	2.01
8	424.18 c	268.48	818.55 b	77.05	2.45 cd	1.29	5.15 bc	1.2	3.68 c	0.78	672.39 bc	404.35	15.35 c	2.08	33.33 cd	3.43	2.51 ab	1.44
9	440.90 bc	334.26	832.50 b	90.70	1.90 cd	0.57	4.38 c	2.13	4.60 c	1.85	461.83 c	174.24	15.79 c	2.42	32.47 cd	1.45	3.10 b	2.91

† Means followed by the same letter within columns are not significantly different for $P < 0.05$ based on LSD of ANOVA.

‡ Number of months with mean max. temperature > 30 °C.

∫ Number of months with mean min. temperature < 7 °C.

¶ Number of months with water deficit measured as the difference between potential and real ETP.

Table 2 Relative frequency (%) of the qualitative agro-morphological traits within the subspecies and populations.

	Subspecies	*dicoccon*		*turgidum*			*durum*						
	Population	1	5	4	5	7	2	3	4	5	6	8	9
	No. genotypes	12	1	21	3	12	19	26	7	9	20	28	22
Growth habit	Prostrate	33.3		61.9	33.3		5.3	3.8	28.6	22.2		14.3	
	Intermediate	66.7	100.0	38.1	66.7	91.7	89.5	96.2	57.1	77.8	45.0	71.4	81.8
	Erect					8.3	5.3		14.3		55.0	14.3	18.2
Awn barbs	Rough	100.0	100.0	100.0	100.0	100.0	94.7	100.0	100.0	100.0	20.0	100.0	81.8
	Smooth						5.3				80.0		18.2
Awnedness	Awnless												4.5
	1- 3 cm			47.6		41.7					10.0		
	3-8 cm	66.7	100.0										
	> 8 cm	33.3		52.4	100.0	58.3	100.0	100.0	100.0	100.0	90.0	100.0	95.5
Spike density	Lax	8.3		4.8					57.1				4.5
	Intermediate	91.7		71.4	33.3	58.3	10.5	15.4	42.9	55.6	60.0	21.4	36.4
	Dense		100.0	14.3	33.3	33.3	63.2	80.8		44.4	40.0	78.6	59.1
	Very dense			9.5	33.3	8.3	26.3	3.8					
Glume hairiness	Hairless	100.0	100.0	57.1	33.3	58.3	21.1	96.2	85.7	66.7	25.0	57.1	72.7
	Low				33.3	8.3	5.3			33.3	15.0	14.3	13.6
	High			42.9	33.3	33.3	73.7	3.8	14.3		60.0	28.6	13.6
Glume colour	White	83.3	100.0	52.4	66.7	50.0	78.9	53.8	28.6	22.2	45.0	42.9	18.2
	Red to brown	16.7		42.9	33.3	41.7	15.8	3.8	71.4	55.6	55.0	46.4	40.9
	Black			4.8		8.3	5.3	42.3		22.2		10.7	40.9
Seed colour	White			9.5		100.0	94.7	100.0	42.9	44.4	90.0	82.1	50.0
	Red	100.0	100.0	90.5	100.0		5.3		57.1	55.6	10.0	17.9	50.0

1 **Table 3** Means and standard deviations (SD) of quantitative agro-morphological traits for each subspecies and population.

	Subspecies	*dicoccon*		*turgidum*			*durum*						
	Population	1	5	4	5	7	2	3	4	5	6	8	9
Days to heading (days)	Mean	180.75	180.00	178.14 a	177.67 a	176.50 a	174.74 c	169.62 a	173.86 bc	171.89 b	172.95 b	173.57 bc	172.09 b
	SD	3.02	-	2.10	2.52	3.09	2.62	2.40	3.93	2.62	2.04	3.24	2.52
Days to maturity (days)	Mean	216.17	214.00	215.62 b	214.00 ab	214.33 a	214.05 bc	211.77 a	215.86 c	214.67 bc	213.40 abc	214.75 c	212.27 ab
	SD	3.43	-	1.50	0.00	1.15	3.75	3.39	1.46	2.18	3.23	3.06	4.85
Grain filling period (days)	Mean	35.42	34.00	37.48 a	36.33 a	37.83 a	39.32 a	42.15 b	42.00 ab	42.78 b	40.45 ab	41.18 ab	40.18 ab
	SD	3.32	-	2.40	2.52	3.13	4.11	3.03	4.12	4.18	2.65	3.54	4.22
Plant height (cm)	Mean	116.67	120.00	136.24 b	124.33 a	130.58 ab	119.37 a	126.04 b	130.57 b	119.22 a	117.90 a	119.25 a	127.95 b
	SD	5.79	-	7.69	4.04	8.11	6.87	6.02	15.38	16.38	7.24	7.09	6.99
Spike length (cm)	Mean	127.75	102.00	106.67 b	88.33 a	116.92 c	89.11 a	89.42 a	119.86 c	97.22 b	95.60 b	97.64 b	96.91 b
	SD	20.15	-	7.49	9.29	6.89	8.60	8.12	15.02	12.56	9.04	6.91	11.33
Spikelets/spike (number)	Mean	26.42	24.00	23.14 a	23.33 ab	28.75 b	22.68 c	22.00 bc	22.29 abc	21.89 bc	21.50 b	23.39 ac	22.27 bc
	SD	3.75	-	1.74	4.51	10.23	1.80	1.17	2.14	1.27	1.73	2.04	1.20
100-grain weight (g)	Mean	12.73	7.52	4.78 a	5.22 ab	5.45 b	5.18 a	5.75 b	5.28 ab	4.84 a	5.89 b	5.17 a	5.14 a
	SD	1.05	-	0.46	0.20	0.54	0.57	0.77	0.90	0.66	0.93	0.66	0.79

2 † Means followed by the same letter within rows and subspecies are not significantly different at $P < 0.05$.

Table 4 Allelic frequencies at each gliadin locus within each subspecies and population.

Locus	Subspecies	*dicoccon*		*turgidum*			*durum*						
	Population	1	5	4	5	7	2	3	4	5	6	8	9
	No. genotypes	12	1	21	3	12	19	26	7	9	20	28	22
Gli-A1	*a*	41.7											
	b		100.0	9.5	33.3	75.0	68.4	7.7	28.6	22.2	45.0	50.0	13.6
	c			42.9			15.8	46.2	28.6	33.3	55.0	3.6	45.5
	e	58.3				16.7	10.5	46.2	42.9	22.2		32.1	13.6
	f						5.3						4.5
	g			19.0	33.3	8.3				11.1		10.7	18.2
	k			14.3	33.3								
	new			14.3						11.1		3.6	4.5
Gli-B1	*a*						10.5	3.8	14.3	33.3		7.1	13.6
	b				33.3	8.3	73.7	3.8				3.6	4.5
	c	50.0		23.8		8.3	15.8	92.3		55.6	50.0	85.7	68.2
	e		100.0										
	new-1	33.3		38.1		50.0			57.1	11.1			9.1
	new-2			14.3									
	new-5					25.0					50.0		
	new-7			14.3	33.3								
	new-8	16.7		9.5		8.3							4.5
	new-9								28.6				
	new				33.3							3.6	
Gli-A2	*a*	41.7		23.8		8.3	10.5			33.3	5.0	3.6	18.2
	b						10.5		14.3		40.0		27.3
	d						5.3			11.1		3.6	
	e			14.3					28.6				
	f	25.0		33.3		41.7						3.6	
	g			4.8			15.8	3.8			25.0	7.1	4.5
	k			19.0		50.0	15.8	3.8	57.1	33.3	25.0	10.7	50.0
	o						5.3	50.0		22.2	5.0	7.1	
	new-1			4.8				7.7				28.6	
	new-2							34.6				3.6	
	new-3						36.8						
	new-4											32.1	
	new	33.3	100.0		100.0								
Gli-B2	*a*									11.1			
	b												4.5
	e												4.5
	h					25.0	47.4	46.2		22.2	40.0	28.6	13.6
	j					8.3							
	l			4.8		8.3	21.1	15.4	28.6		45.0	25.0	36.4
	o										5.0		
	t			4.8				38.5	14.3	22.2		21.4	4.5
	new-1			14.3	33.3	33.3	21.1		14.3	11.1			
	new-2			14.3									
	new-3												9.1
	new-4			9.5			5.3					7.1	4.5
	new-5			9.5			5.3					3.6	4.5
	new	100.0	100.0	42.9	66.7	25.0			42.9	33.3	10.0	14.3	18.2

Table 5 F_{ST} values assessed with SSR and gliadins.

	SSR	Gliadins
Among subspecies	0.12**	0.11**
dicoccon vs *turgidum*	0.18**	0.17**
dicoccon vs *durum*	0.12**	0.14**
turgidum vs *durum*	0.10**	0.09**
Among agro-ecological zones	0.12**	0.06**
Among populations	0.20**	0.18**

** Significant at the 0.01 probability level.

Table 6 F_{ST} values between populations (Pop) assessed with SSR (above diagonal) and gliadins (below diagonal).

Pop.	1	2	3	4	5	6	7	8	9
1	-	0.35**	0.35**	0.15**	0.13**	0.29**	0.21**	0.22**	0.25**
2	0.34**	-	0.28**	0.23**	0.19**	0.31**	0.27**	0.20**	0.19**
3	0.31**	0.33**	-	0.24**	0.20**	0.27**	0.31**	0.20**	0.20**
4	0.16**	0.20**	0.24**	-	0.06**	0.20**	0.12**	0.15**	0.17**
5	0.15**	0.12**	0.16**	0.03	-	0.15**	0.11**	0.07**	0.11**
6	0.33**	0.20**	0.24**	0.16**	0.13**	-	0.25**	0.19**	0.19**
7	0.27**	0.17**	0.37**	0.09*	0.10	0.21**	-	0.18**	0.21**
8	0.24**	0.21**	0.14**	0.18**	0.08**	0.17**	0.22**	-	0.12**
9	0.23**	0.21**	0.18**	0.08**	0.04	0.08**	0.20**	0.12**	-

* Significant at the 0.05 probability level.

** Significant at the 0.01 probability level.

Table 7 Distribution (%) of the subspecies and populations within each of the nine agro-ecological zones (Z).

Subspecies	*dicoccon*		*turgidum*			*durum*							
Population	1	5	4	5	7	2	3	4	5	6	8	9	Total
No. genotypes	12	1	21	3	12	19	23	6	9	20	27	22	180
Z1	16.7		28.6		8.3	5.3		50.0		10.0	3.7		16
Z2				33.3	16.7	10.5	30.4		22.2	5.0	25.9	4.5	23
Z3					8.3	31.6	17.4			10.0	22.2	4.5	20
Z4	16.7	100.0	14.3		25.0	15.8	8.7	50.0	22.2	5.0	7.4	22.7	27
Z5					33.3	15.8	4.3		11.1		29.6	4.5	18
Z6			9.5			21.1	17.4			10.0	3.7	40.9	22
Z7							13.0		22.2		3.7		6
Z8			9.5		8.3		8.7		22.2	60.0		22.7	24
Z9	66.7		38.1	66.7							3.7		19
unknown													5

Table S1: Chromosomal location, number of alleles (unique alleles), major allele frequency and genetic diversity (PIC) of the 39 microsatellite loci analyzed.

SSR Marker	Chromosome	Allele number	Major Allele Frequency	PIC
BARC1032	5B	8 (5)	0.81	0.33
BARC1077	3B	4 (1)	0.59	0.54
BARC155	5A	11 (4)	0.42	0.67
BARC55	2B	10 (4)	0.33	0.74
BARC80	1B	17 (6)	0.21	0.88
CFA2219	1A	19 (8)	0.19	0.88
CFA2257	7A	15 (3)	0.62	0.58
CFA2263	2A	14 (9)	0.37	0.75
WMC468	4A	5 (1)	0.65	0.50
WMC522	2A	32 (8)	0.11	0.94
Xgwm0002	3A	4 (0)	0.79	0.36
Xgwm0011	1B	15 (3)	0.24	0.88
Xgwm0018	1B	12 (4)	0.39	0.72
Xgwm0046	7B	15 (4)	0.22	0.88
Xgwm0060	7A	19 (5)	0.33	0.80
Xgwm0088	6B	17 (4)	0.20	0.72
Xgwm0095	2A	6 (0)	0.37	0.80
Xgwm0099	1A	11 (3)	0.30	0.94
Xgwm0136	1A	41 (14)	0.15	0.78
Xgwm0148	2B	12 (4)	0.32	0.80
Xgwm0154	5A	16 (6)	0.30	0.85
Xgwm0155	3A	16 (3)	0.24	0.86
Xgwm0156	5A	20 (4)	0.27	0.91
Xgwm0181	3B	22 (5)	0.15	0.80
Xgwm0186	5A	28 (15)	0.35	0.77
Xgwm0234	5B	19 (9)	0.39	0.75
Xgwm0251	4B	11 (1)	0.42	0.86
Xgwm0299	3B	21 (7)	0.30	0.83
Xgwm0312	2A	26 (4)	0.45	0.85
Xgwm0332	7A	20 (4)	0.27	0.86
Xgwm0389	3B	14 (3)	0.22	0.75
Xgwm0408	5B	17 (8)	0.39	0.59
Xgwm0459	6A	23 (8)	0.26	0.81
Xgwm0494	3A	9 (2)	0.38	0.95
Xgwm0513	4B	4 (0)	0.59	0.79
Xgwm0570	6A	15 (3)	0.33	0.90
Xgwm0577	7B	41 (15)	0.10	0.77
Xgwm0601	4A	14 (4)	0.34	0.75
Xgwm0604	5B	18 (6)	0.33	0.81
Mean		16,44 (5)	0.35	0.77

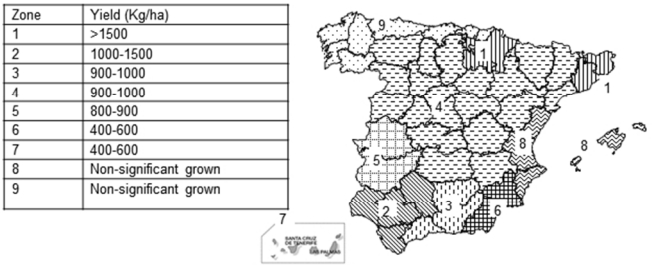


Figure 1. Map of Spanish provinces showing the nine agro-ecological zones for durum wheat cultivation according to long-term (1920-1991) average yield at each province.
254x190mm (96 x 96 DPI)

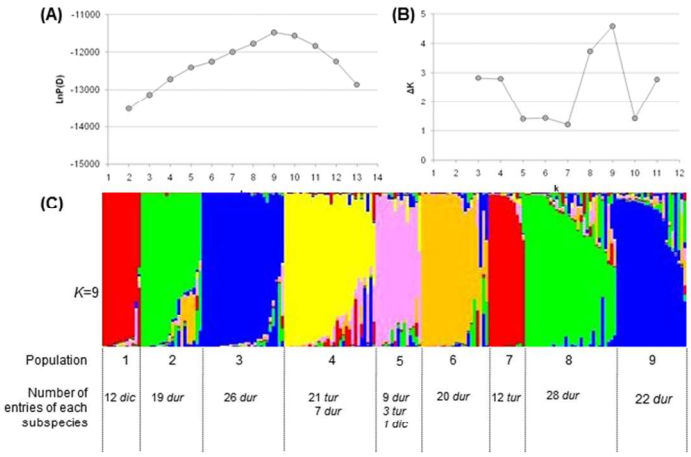


Figure 2. (A) Changes in the natural log probability of the data against the number of subpopulations (k). (B) Average Δk against k. Values are the mean of 5 runs of STRUCTURE simulations. (C) Inferred structure of the durum wheat collection based on 190 genotypes and genomic SSR markers using STRUCTURE. Each individual is represented by a line divided into colored segments that represent the individual's estimated membership fractions to each of the nine subpopulations. The subspecies are indicated as dic = dicoccon, dur = durum and tur = turgidum.
254x190mm (96 x 96 DPI)

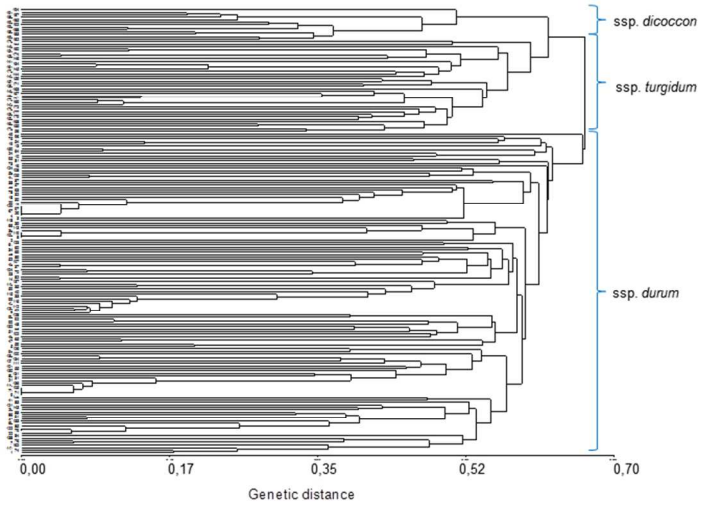


Figure 3. Dendrogram of the DArT genotypes of the 184 accessions analyzed.
254x190mm (96 x 96 DPI)

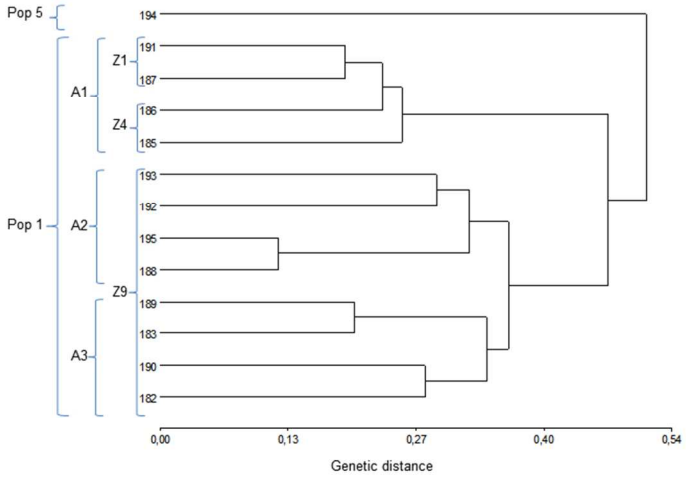


Figure 4. Dendrogram of the DArT genotypes of the ssp. dicoccon. The populations (Pop), DArT subgroup (A) and agro-ecological zone (Z) are indicated.
254x190mm (96 x 96 DPI)

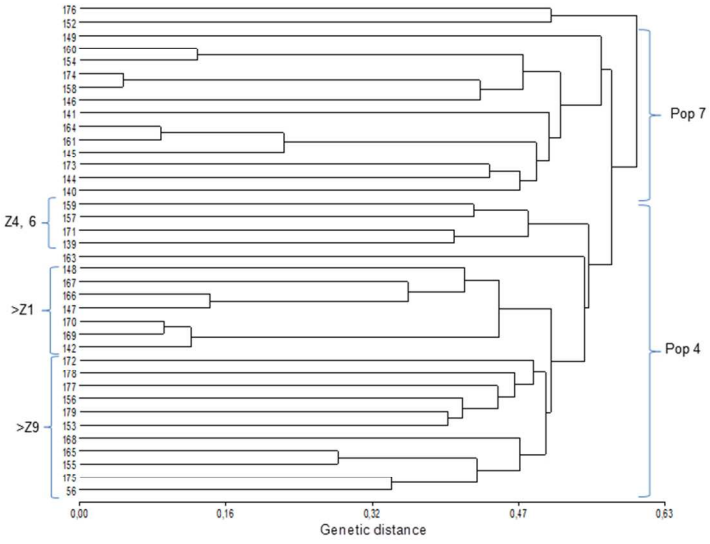
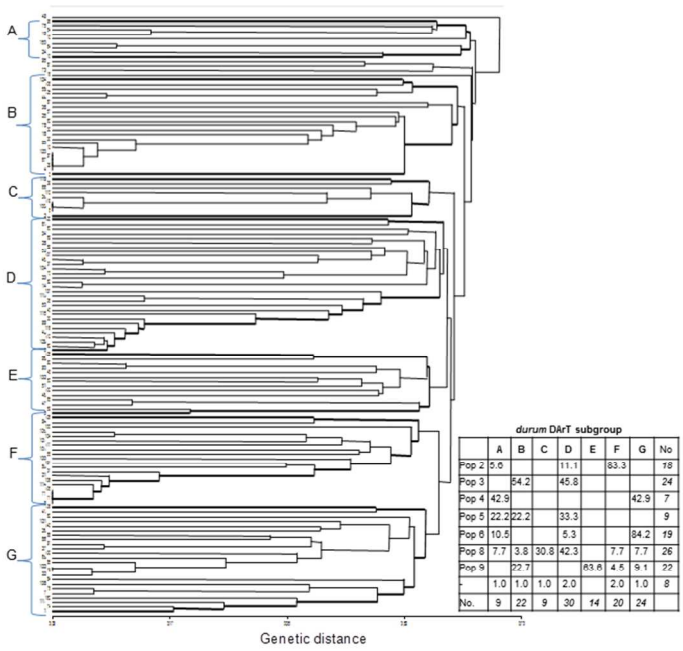


Figure 5. Dendrogram of the DArT genotypes of the spp. turgidum. The populations (Pop) and agro-ecological zone (Z) are indicated.
254x190mm (96 x 96 DPI)



	A	B	C	D	E	F	G	No
Pop 2	5.6			11.1		83.3		18
Pop 3		54.2		45.8				24
Pop 4	42.9						42.9	7
Pop 5	22.2	22.2		33.3				9
Pop 6	10.5			5.3			84.2	19
Pop 8	7.7	3.8	30.8	42.3		7.7	7.7	26
Pop 9		22.7			63.6	4.5	9.1	22
-	1.0	1.0	1.0	2.0		2.0	1.0	8
No.	9	22	9	30	14	20	24	

Figure 6. Dendrogram of the DArT genotypes of the ssp. durum. The DArT subgroups (A-G) are outlined in bold. The table shows the percentage of each population (Pop) classified in a DArT subgroup.
254x190mm (96 x 96 DPI)