Field Crops Research 119 (2010) 91–105

Contents lists available at ScienceDirect

Field Crops Research

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

Understanding the relationships between genetic and phenotypic structures of a collection of elite durum wheat accessions

Conxita Royo[a,*], Marco Maccaferri[b], Fanny Álvaro[a], Marc Moragues[a], Maria Corinna Sanguineti[b], Roberto Tuberosa[b], Fouad Maalouf[c], Luis F. García del Moral[d], Andrea Demontis[e], Sayar Rhouma[f], Miloudi Nachit[g], Nasserlehaq Nserallah[h], Dolors Villegas[a]

[a] IRTA (Institute for Food and Agricultural Research and Technology), Field Crops Section, Cereal Breeding, Centre UdL-IRTA, Rovira Roure, 191, 25198 Lleida, Spain
[b] Department of Agroenvironmental Sciences and Tecnology, University of Bologna, 40127 Bologna, Italy
[c] Department of Plant Breeding, Lebanese Agricultural Research Institute, Zahleh, Lebanon
[d] Department of Plant Physiology, University of Granada, 18071 Granada, Spain
[e] Società Produttori Sementi Bologna, Research Division, 40050 Argelato, Italy
[f] Tunisian National Institute of Agronomic Research, 2080 Tuni, Tunisia
[g] ICARDA, Aleppo, Syria
[h] CRRA-INRA, Settat, Morocco

ARTICLE INFO

Article history:
Received 26 October 2009
Received in revised form 29 June 2010
Accepted 29 June 2010

Keywords:
Triticum turgidum
Germplasm collection
Genetic diversity
Genotype × environment interaction
DARWin

ABSTRACT

A collection of 191 durum wheat accessions representing Mediterranean Basin genetic diversity was grown in nine different environments in four countries, with productivities ranging from 0.99 to 6.78 t ha^{-1}. The population breeding structure comprised eight genetic subpopulations (GSPs) using data derived from 97 evenly distributed SSR markers. The phenotypic structure was assessed: (i) from the mean values of six agronomic traits across environments (multivariate), and (ii) from data representing each trait in each environment (univariate). Mean daily maximum temperature from emergence to heading was significantly ($P<0.05$) and negatively associated to yield, accounting for 59% of yield variations. Significant but weak relationships were obtained between the genetic similarities among accessions and their overall agronomic performance ($r=0.15$, $P<0.001$), plant height ($r=0.12$, $P<0.001$), spike–peduncle length ($r=0.06$, $P<0.01$) and thousand kernel weight ($r=0.03$, $P<0.05$), suggesting a very low possibility of prediction of the agronomic performance based on random SSR markers. The percentage of variability (measured by sum of squares) explained by the environment varied between 76.3 and 98.5% depending on the trait, while that explained by genotypes ranged between 0.4 and 12.6%, and that explained by the GE interaction ranged from 1.1 to 12.5%. The clustering of the accessions based on multivariate phenotypic data offered the best explanation of genotypic differences, accounting for 30.3% (for yield) to 75.1% (for kernel weight) of the observed variation. The genotype × environment interaction was best explained by the phenotypic univariate clustering procedure, which explained from 28.5% (for kernel weight) to 74.9% (for days to heading) of variation. The only accessions that clustered both in the genetic dissimilarities tree and the tree obtained using Euclidean distances based on standardized phenotypic data across environments were those closely related to the CIMMYT hallmark founder 'Altar 84', the ICARDA accessions adapted to continental-dryland areas, and the landraces, suggesting that genetic proximity corresponded to agronomic performance in only a few cases.

1. Introduction

Durum wheat (*Triticum turgidum* L. var *durum*) is a commercially important tetraploid wheat species that originated and diversified in the Mediterranean Basin (Mac Key, 2005). This region today accounts for ca. 60% of global production and represents the greatest source of genetic diversity in durum wheat germplasm, as indicated by the high levels of polymorphism present in the elite accessions (Maccaferri et al., 2003) and in local landraces (Moragues et al., 2006, 2007). The Mediterranean Basin is also characterized by highly variable environments. Although most durum wheat in the region is grown under rain-fed conditions, rainfall pattern is rather unpredictable and drought is an important yield constraint (Araus et al., 2003). This combination of genetic and environmental diversity results in large spatial and temporal yield fluctuations.

In several Mediterranean countries, breeding programs have attempted to generate varieties that produce high yields despite the

* Corresponding author. Tel.: +34 973 702583; fax: +34 973 238301.
E-mail address: conxita.royo@irta.es (C. Royo).

0378-4290/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.fcr.2010.06.020

variable climate, by combining the advantageous adaptive traits of landraces and old cultivars with the high yield potential of modern varieties. Such programs have been developed in Italy (Boggini et al., 1990), Tunisia (Daaloul et al., 1990) and Syria (Srivastava, 1987). Two CGIAR centers, CIMMYT (International Maize and Wheat Improvement Center) and ICARDA (International Center for Agricultural Research in Dry Areas), have facilitated the distribution and exchange of durum wheat germplasm through their interactions with the National Breeding Programs in the region.

Greater knowledge of the genetic relationships among accessions of different origin facilitates their classification and characterization, leading to the definition of crosses and selection strategies. A number of methods have been used to analyze the genetic diversity of crops, typically utilizing pedigree records (van Hintum and Haalman, 1994), agronomic and morphological data (Jain et al., 1975; Porceddu, 1976; Peccetti and Annicchiarico, 1993), seed storage proteins (van Hintum and Elings, 1991; Moragues et al., 2006) and, more recently, DNA markers (Autrique et al., 1996; Soleimani et al., 2002). Molecular markers have been successfully used in durum wheat to determine genetic relationships and population structure (Maccaferri et al., 2005; Moragues et al., 2007). Such markers are advantageous because they are not influenced by the environmental factors that affect agronomic traits. Simple sequence repeats (SSRs, microsatellites) are excellent markers for genetic diversity analysis and genotyping in crop species such as wheat (Donini et al., 2000), because they are widely distributed in the genome, codominant, highly polymorphic, stable, reproducible and relatively simple to analyze (Fufa et al., 2005; Song et al., 2005). Genetic similarities based on SSR data and the model-based cluster analysis implemented in the software STRUCTURE (Pritchard et al., 2000; Falush et al., 2003) have been used to identify subpopulations within wheat collections (Maccaferri et al., 2005, 2006; Somers et al., 2007).

So far, only few studies have examined the relationship between genetic population structure and field performance in wheat. The studies conducted by Fufa et al. (2005) using 30 bread wheat genotypes and by Annicchiarico et al. (2009) using 24 durum wheat genotypes revealed little correlation between phenotypic traits and genetic distance estimates from molecular markers; however, both these studies involved rather small set of accessions. The hypotheses that we wanted to test in this study were (i) when using a large number of genotypes, similarities for molecular markers should provide some indications on similarities for adaptive response, and (ii) the clustering of accessions based on agronomic traits' data should be more informative to explain the genotype (G) effect and the genotype × environment (GE) interaction of the ANOVA than the clustering based on random molecular markers. The objectives of the study were (i) to analyze the relationship between the genetic and phenotypic structures of a collection of 191 durum wheat accessions of diverse origin assembled and evaluated in the framework of the EU funded project IDuWUE (http://www.distagenomics.unibo.it/iduwue/index.html), (ii) to identify the clustering procedure that better explained the G effect and the GE interaction for a set of agronomic traits, and (iii) to perform diversity analyses on the basis of molecular and phenotypic dissimilarities between accessions. An additional objective was to identify the environmental traits mostly affecting durum wheat yield under Mediterranean conditions.

2. Materials and methods

2.1. Characterization of environments

Maximum and minimum temperatures and water input were measured daily in weather stations close to the experimental fields. Reference evapotranspiration (ET_0) was computed daily according to the FAO-56 guidelines (Allen et al., 1998). The environmental variables considered from emergence to heading and from heading to two weeks after heading were, respectively: mean of daily minimum and maximum temperatures, water input and reference evapotranspiration. Reference evapotranspiration was also calculated from emergence to harvesting. Soil cores from a depth of 15 cm were taken every other week from the beginning of jointing to heading, and every week from heading to maturity. Soil moisture in the cores was determined using the gravimetric method (Campbell and Mulla, 1990). The area under the curve (AUC) of the percentage of soil moisture was estimated as:

$$\mathrm{AUC} = \sum_{i=1}^{n-1} (A_i + A_{i+1 \ldots} A_{n-1}),$$

with $A_i = \min(m_i, m_{i+1}) \times (d_{i+1} - d_i) + [|m_i - m_{i+1}| \times (d_{i+1} - d_i)]/2$,

where n is the number of moisture sampling dates, min denotes the minimum value between brackets, m_i is the soil moisture measured on day i, and d_i is the number of days from sowing to day i. The AUC of the percentage of soil moisture at 15 cm depth was calculated from one week before heading to three weeks after heading, from jointing to heading, and from heading to maturity.

2.2. Plant material

A collection of 191 elite durum wheat accessions was assembled in order to sample a large portion of the genetic diversity of durums cultivated in the Mediterranean Basin. These accessions were chosen from a larger collection comprising 330 accessions of different origin and evaluated in a comparative field trial carried out in 2003 in Cadriano, Italy (Maccaferri et al. unpublished). Accessions were chosen according to their relatedness and phenological uniformity (especially as to heading date). According to the country/institution of origin, the accessions were grouped as follows: (i) 83 from ICARDA; (ii) 39 from Italy; (iii) 24 from Spain; (iv) 18 from CIMMYT, some of which have been used in National Breeding Programs throughout the Mediterranean Basin; (v) 13 from Morocco; (vi) two from Tunisia; (vii) 10 from the germplasm cultivated under irrigation in the south-western region of the USA (Arizona/California), and commonly referred to as 'desert durums'; and (viii) the widely grown cultivars Simeto (from Italy) and Vitron (from Spain) as reference checks. Most of the accessions were semi-dwarf elite materials released from the early 1970s until the late 1990s. The collection also included a limited number of important 'founder genotypes' widely used as parents in breeding programs throughout the Mediterranean Basin. Seeds of the accessions and checks were increased in a single location (Cadriano, Bologna, 2003).

2.3. Molecular profiling

For each accession, genomic DNA was extracted from freeze-dried young leaves pooled from 20 seedlings per accession according to Saghai Maroof et al. (1984). The SSR markers for genotyping were chosen among those that were polymorphic in two durum recombinant inbred line (RIL) populations used in previous SSR mapping studies (Maccaferri et al., 2008; Mantovani et al., 2008) and the genetic structure of the population was estimated using 97 SSRs (for details see Maccaferri et al., 2006).

2.4. Field experiments and phenotypic data

The 191 accessions were evaluated during the 2003–2004 and 2004–2005 growing seasons in nine field experiments in four countries (acronyms and details of experiments are reported in Table 1).

Table 1
Experimental details of the nine field trials.

Environment[a]	Itl1-r04	Itl2-r05	Lbn-i04	Lbn-r04	Lbn-r05	Spn1-r04	Spn2-r05	Tns-i05	Tns-r05
Environment number	1	2	3	4	5	6	7	8	9
Country	Italy	Italy	Lebanon	Lebanon	Lebanon	Spain	Spain	Tunisia	Tunisia
Site	Cadriano	Cerignola	Rayack	Rayack	Rayack	Gimenells	Granada	Kef	Kef
Year	2004	2005	2004	2004	2005	2004	2005	2005	2005
Coordinates	44° 33′ N 11° 24′ E	41° 28′ N 15° 84′ E	33° 51′ N 35° 59′ E	33° 51′ N 35° 59′ E	33° 51′ N 35° 59′ E	38° 56′ N 0° 29′ E	37° 15′ N 3° 46′ W	36° 14′ N 8° 27′ E	36° 14′ N 8° 27′ E
Seeding date	11/06/2003	23/11/2004	21/11/2003	21/11/2003	12/11/2004	16/12/2003	12/10/2004	28/11/2004	28/11/2004
Harvest date	07/09/2004	14/06/2005	06/05/2004	06/05/2004	20/06/2005	13/07/2004	07/08/2005	31/05/2005	31/05/2005
Water input (mm)									
Emergence–heading	526	195	611	611	369	195	175	277	227
Heading–two weeks after heading	6	3	99	5	8	0	4	36	36
Emergence–harvesting	584	207	711	616	378	250	187	317	267
Mean daily minimum temperature (°C)									
Emergence–heading	3	5	3	3	3	4	1	4	4
Heading–two weeks after heading	10	13	8	8	5	8	11	8	8
Mean daily maximum temperature (°C)									
Emergence–heading	11	14	15	15	16	14	17	13	13
Heading–two weeks after heading	23	27	20	20	22	25	29	19	19
Reference evapotranspiration (mm)									
Emergence–heading	201	411	324	319	345	180	263	206	206
Heading–two weeks after heading	67	89	66	64	67	31	72	50	50
Emergence–harvesting	489	733	633	633	672	465	736	383	383
AUC of the percentage of soil moisture at 15 cm depth (%day)									
Jointing–heading	14.44	7.08	8.95	6.78	4.78	10.75	2.46	7.40	3.48
1 week before heading–3 weeks after heading	8.48	4.61	4.00	3.55	4.02	3.85	3.08	6.54	3.61
Heading–maturity	9.02	5.49	4.71	4.42	5.23	3.33	5.17	7.75	5.37
Mean yield (kg ha^{-1})	6476	5662	4090	2213	3598	6453	993	6780	4377
Yield intrablock $\hat{\sigma}^2_{error}$	0.043	0.0122	0.036	0.0207	0.0179	0.0167	0.0012	0.041	0.018

[a] *r* and *i* after the dash indicate rainfed or irrigated site, respectively.

Field experiments consisted of non-replicated plots (4 m^2, comprising eight 2.5-m rows, spaced 0.20 m apart) arranged according to a modified augmented design with three checks, two of which (cultivars Simeto and Vitron) were common to all experiments. The additional check was chosen among the best locally adapted varieties. This design was chosen because it allowed testing such a large number of entries saving land and management costs, having the additional advantages of being independent of the homogeneity of error variance and of the model of response at each environment (Federer et al., 2001). Seed density was adjusted to 400 germinable seeds/m^2. In order to prevent attacks from seed-transmitted fungal diseases, seed was treated with Vitavax FLO NF (Carboxin + Thiram). Agronomic management (including fertilization and pest, disease and weed control) was carried out according to the standard agriculture practices in each site and country.

The mean date of plant emergence was recorded in each experiment. Heading date was recorded when more than 50% of the main spikes within a plot had reached Zadoks stage 55 (Zadoks et al., 1974). Grain yield was determined by mechanically harvesting the plots at ripening, and was expressed at a 12% moisture level. Plant height and peduncle length were measured during grain filling on three main stems randomly taken per plot considering the distance from the ground to the tip of the ear without awns and the distance from the flag leaf collar to the base of the ear, respectively. The number of grains per spike was determined at maturity from a sample of the spikes contained in one linear meter on a central row per plot. Thousand kernel weight (TKW) was measured on two independent 100 seeds samples per plot. No lodging was detected in all the experiments.

2.5. Statistical analyses

2.5.1. Environmental

To identify the environmental variables that mostly affected yield stepwise regression analysis was carried out with the mean yield of each environment as a dependent variable and with all the environmental variables described in Section 2 as independent variables.

2.5.2. Genetic structure

The genetic structure of the collection was determined with Bayesian methods using the STRUCTURE software (Pritchard et al., 2000), with the optimum number of subpopulations equal to eight ($K = 8$; Maccaferri et al., 2006).

2.5.3. Phenotypic structures

Phenotypic data were fitted to a linear mixed model with the check cultivars as fixed effects, and the row number, column number and accession as random effects (Little et al., 1996). Restricted Maximum Likelihood (REML) was used to estimate the variance components and to produce the Best Linear Unbiased Predictors (BLUPs) for the phenotypic data of each accession within each environment, achieved using the MIXED procedure of the SAS-STAT statistical package (SAS Institute Inc., 2000).

The phenotypic structures of the population were assessed according to both multivariate (considering the six phenotypic traits studied) and univariate procedures. In the last, a different clustering was obtained for each trait. Multivariate clustering was performed by applying the furthest neighbor algorithm to the Euclidean distance matrix, calculated from the mean values of the BLUPs of the six traits across environments, using the GENSTAT (GenStat 11th edition, http://www.genstat.com) software. Clustering of accessions based on univariate phenotypic data was assessed using the Corsten and Denis algorithm (Corsten and Denis, 1990), which grouped accessions and environments that behaved similarly for each particular trait. The CINTERACTION procedure in GENSTAT was applied to the data matrix representing values for each trait in each accession in each environment (Romagosa et al., 2009). Given that the intra-block errors underestimated the real value, the errors for stopping the clustering procedures were estimated through the non-significant eigenvalues of an AMMI analysis applied to the table of means. Errors were computed as the quotient between the total sum of squares of the non-significant eigenvalues and the sum of their degrees of freedom. Eigenvalues were considered significant when they explained >30% above the average percentage explained by an eigenvalue.

2.5.4. Associations between matrices based on genetic and phenotypic data

The Mantel test (Mantel, 1967), which is based on product–moment correlation, was used to calculate the associations between the similarity matrix of genetic distances and the matrices of distances calculated from multivariate and univariate phenotypic data. The matrix of genetic distances was calculated using the Manhattan coefficient on a matrix of probabilities that each accession belonged to each genetic subpopulation according to SSR data and the STRUCTURE software. The matrix of Euclidean distances based on the data set of the six phenotypic variables (multivariate) was calculated from the mean data for each accession across environments. Binary similarity matrices (1 when two accessions belonged to the same subpopulation and 0 when they belonged to different ones) were created for each phenotypic trait applying the Simple Matching test to the matrices containing the population structure obtained using the Corsten and Denis algorithm. To assess the significance of the associations, 5000 permutations were carried out in which the rows/columns of the phenotypic matrices were randomly permuted. The significance of the associations was estimated by the percentage of the random permutations whose association was greater or equal to that of the original matrices. Calculations were made using the GENSTAT software.

2.5.5. Analyses of variance

Analyses of variance (ANOVA) were performed for each phenotypic trait considering the genotype (G), environment (E) and GE interaction as sources of variation. In an attempt to detect the clustering procedure that better explained the observed variation for each trait, the sum of squares of the G and GE effects was partitioned in the ANOVA according to the subpopulations revealed by genetic (structured and unstructured accessions), phenotypic multivariate and phenotypic univariate analysis. Means were compared using Duncan's multiple range test at $P = 0.05$.

2.5.6. Diversity analyses

Diversity analysis was conducted using both molecular and phenotypic data. Genetic relationships among accessions were determined by means of Dice's coefficient (Dice, 1945) from the binary matrix of 72 SSRs with known chromosome positions. Phenotypic relationships were determined from the Euclidean distances calculated with the standardized mean phenotypic data across environments. Un-rooted trees were constructed using the hierarchical clustering method of the software package DARWin5.0 (Perrier et al., 2003).

3. Results

3.1. Environmental data

The nine environments represented a broad range of growing conditions across the Mediterranean Basin. Experiments were conducted at latitudes from 33°51′ N to 44°33′ N and longitudes from 35°59′ E to 3°46′ W (Table 1). Water input (rainfall plus supplementary irrigation) ranged from 187 to 711 mm during the entire crop cycle. The sowing-to-heading period was within a 101–172 d range

and average environmental yields ranged from 0.99 to 6.78 t ha^{-1} (Table 1).

To explore the influence of the environmental factors on grain yield, regression analysis was carried out with the environmental data as independent variables and yield as dependent variable. The only variable that entered in the model ($P < 0.05$) was the maximum temperature from emergence to heading, which was negatively associated with yield ($R^2 = 0.59$, $P < 0.05$).

3.2. Clustering of accessions

3.2.1. Genetic structure

The genetic structure of the population was described by eight genetic subpopulations (GSPs) derived from SSR markers used to characterize the 191 accessions (Maccaferri et al., 2006). Using the cut off $P \geq 0.50$ only 113 accessions were assigned to one GSP (Table 2a), whereas 78 accessions had <50% likelihood of belonging to any GSP and thus were considered as unstructured (Table 2b).

The analysis of the genetic classification of the accessions assigned to each GSP together with their pedigree and/or origin (Table 2a) allowed us to make the following considerations. GSPs 1 and 2 included accessions closely related to the CIMMYT hallmark founders 'Altar 84' (selected from the cross Ruff's'/Flamingo's'//Mexicali75/3/Shwa's') and 'Yavaros79' (developed from 'Bittern' with pedigree Jori's'//Anhinga's'/Flamingo's'). 'Altar84' is characterized by high yield potential and 'Yavaros79' by wide adaptation. GSPs 3, 4 and 5 contained sets of accessions bred at ICARDA for adaptation to specific environments: high-yield (GSP

Table 2a
List of the accessions used in the study. The 113 accessions assigned to a subpopulation (structured), i.e. with a probability higher than 50% of belonging to any of the eight genetic subpopulations, as identified based on simple sequence repeat (SSR) markers.

Subpopulation 1 CIMMYT ('Altar 84')			Subpopulation 3 ICARDA (High yield)			Subpopulation 5 ICARDA (continental-dryland)			Subpopulation 7 Spanish		
	O[a]	%[b]		O	%		O	%		O	%
Ahi-ou-1	C	97	Aghrass-1	IC	94	Massara-1	IC	95	Boabdil	S	70
Gs/Cra//Sba81/3/Ho	C	95	Terbol97-3	IC	90	Omrabi-5	IC	92	Bolido	S	70
Kulrengi-Balikcil_8	C	95	Bcrch-1	IC	88	Omrabi-3	IC	91	Artena	S	69
Gallareta	S	93	Amedakul-1	IC	84	Tomouh	IC	88	Roqueño	S	68
Acuatico/Yazi_1	C	92	Loukos-1	IC	77	Ombit-1	IC	87	Dukem/3/Ruff/Fgo...	C	57
Focha_1/5*Alas	C	88	Ainzen-1	IC	75	Mrb-17	IC	84	Sebou	IC	55
Yazi-10-1	C	88	Miki-1	IC	75	Aw12/Bit	IC	82	Durcal	S	52
Rok/Fgo//Stil/3/Bisu_1	C	86	Ammar-1	IC	74	Younes-1	IC	73	Total accession (no.)	7	
Bushen_4/Tarro_2...	C	86	Bicrederaa-1	IC	72	Omlahn-3	IC	72			
Sooty_9/2*Tarro_1	C	85	Ouasloukos-1	IC	70	Capeiti-8	IT	69			
Rascon_37/2*Tarro_2	C	84	Bicre	IC	69	Omsnima-1	IC	68			
Plata-16	C	83	Bic/3/Cham1...	IC	68	Blk2//134xS-69...	IC	64			
Bombasi	S	83	Arislahn-5	IC	59	Platani	IT	56			
Porto-5	C	82	Cham-1	IC	58	Total accessions (no.)	13				
Topdy_21/Rascon_33	C	81	Maamouri-1	IC	55						
Sula	S	75	Azeghar-2	IC	54						
Bisu_1/Patka_3	C	75	Osa-1/Stj-5.	IC	54						
Arcobaleno	S	65	Total accessions (no.)	17							
Illora	S	64									
Astigi	S	64									
Marjana	M	50									
Total accessions (no.)	21										

Subpopulation 2 CIMMYT ('Yavaros79')			Subpopulation 4 Moroccan + ICARDA Temperate dryland			Subpopulation 6 Italian ('Valnova')			Subpopulation 8 Landrace-derived		
	O	%		O	%		O	%		O	%
Karim	T	96	Isly	M	89	Valnova	IT	92	Haurani	IC	87
Yasmine	M	95	Morocco1807	M	85	Anton	S	90	Shahba	IC	85
Duilio	IT	94	Morocco1808	M	80	Bradano	IT	88	Aldeano	S	78
Vitron	S	94	Messapia	IT	76	Bravadur	U	83	Saada3/Dds//Mtl1	IC	63
Ourgh	M	94	Morocco1804	M	76	Durex	U	83	Valbelice	IT	61
Anouar	M	93	Quadalete	IC	76	Gargano	IT	81	Ouassel-1/4/Buc...	IC	55
Jawhar	M	75	Furat-1	IC	71	Ofanto	IT	78	Trinakria	IT	50
Bronte	IT	69	Produra	IT	61	Colorado	U	77	Total accessions (no.)	7	
Morocco1805	M	64	H.Moul/Chaba88	IC	60	Ixos	IT	77			
Canyon	S	56	Marzak	M	58	Varano	IT	76			
Borli	S	56	Morocco1809	M	55	Simeto	IT	74			
Tunsyr-1	IC	54	Total accessions (no.)	11		Fortore	IT	73			
Meridiano	IT	53				West Bred-881	U	73			
Tensift-1	IC	51				Mexicali-75	C	73			
Total accessions (no.)	14					Grazia	IT	70			
						Kronos	U	69			
						Cannizzo	IT	68			
						Mongibello	IT	65			
						Quadrato	IT	57			
						Torrebianca	IT	56			
						Kofa	U	53			
						Reva	U	52			
						Plinio	IT	51			
						Total accessions (no.)	23				

[a] O = Origin of the accessions according to the following code: C = CIMMYT, IC = ICARDA, IT = Italy, M = Morocco, S = Spain, T = Tunisia, U = USA.
[b] Probability of belonging to the subpopulation.

Table 2b
List of the accessions used in the study. The 78 accessions unassigned to any subpopulation (unstructured), i.e. with less than 50% probability of belonging to any of the eight genetic subpopulations.

Name	O[a]	Name	O	Name	O	Name	O
Angre	S	Claudio	IT	Khabur-1	IC	Ouaslahn-1	IC
Appio	IT	CMH82A.1062...	C	Krf	IC	Pietrafita	IT
Appulo	IT	Colosseo	IT	Krs/Hau	IC	Quabrach-1	IC
Arcangelo	IT	Cortez	U	Lagonil-2	IC	Quad//Erp/Mal...	IC
Aric 31708.70/3/Bo...	IC	Creso	IT	Lagost-3	IC	Radioso	IT
Ariesol	S	Deraa	IC	Lahn	IC	Razzak	T
Atlast-1	IC	Don Pedro	S	Lesina	IT	Sajur	IC
Aus-1	IC	Duroi	S	Lira B-45	IT	Sebah	IC
Awali-1	IC	Flaminio	IT	Maryr-1	IC	Sebatel-1	IC
Bigost-1	IC	Geromtel-1	IC	Mohawk	U	Senadur	S
Blk-2	IC	Gezira-17	IC	Mousabil-2	IC	Stojocri-3	IC
Bolenga	S	Gidara-2	IC	Murlagost-1	IC	Svevo	IT
Bolo	S	Gr/Boy	IC	Nile	IC	Tarek	M
Brachoua	IC	Guerou-1	IC	Norba	IT	Telset-5	IC
C266	C	Heider	IC	Ombar	IC	Wadalmez-1	IC
Cappelli	IT	Iride	IT	Omgenil-3	IC	West Bred Turbo	U
Chaba/Deraa	IC	Italo	IT	Omruf-2	IC	Yousef-1	IC
Chabha-88	IC	Jabato	S	Ort-1	IC	Zeina-1	IC
Chacan	IC	Jordan	IC	Otb-6	IC		
Ciccio	IT	Kabir-1	IC	Ouaserl-1	IC	Number of accessions	78

[a] O = Origin of the accessions according to the following code: C = CIMMYT, IC = ICARDA, IT = Italy, M = Morocco, S = Spain, T = Tunisia, U = USA.

3), temperate-dryland (GSP 4), and continental-dryland areas (GSP 5, Nachit, personal communication). The founders 'Omrabi' and 'Capeiti 8', bred at ICARDA and Italy from the *syriacum* durum types 'Haurani' and 'Eiti', respectively, were included in GSP 5, whereas most Moroccan accessions were placed in GSP 4. Most GSP 6 accessions included the Italian cultivar 'Valnova' in their pedigree, such as the check cultivar 'Simeto' (pedigree: Capeiti 8/Valnova). Five of the seven accessions included in GSP 7 were Spanish. Finally, GSP 8 mostly comprised landrace-derived genotypes of diverse origin.

Mean phenotypic values across environments of the eight GSPs are shown in Table 3. The mean yield of the GSPs ranged from 4.18 to 4.54 t ha^{-1}. GSPs 1 and 2 were the most productive, whereas GSP 8, representing landraces, had the lowest yield. GSP 1 had the highest number of grains per spike and GSP 8 the lowest. The heaviest grains were recorded in GSPs 6 and 2, and the lightest in GSP 1. The largest difference between GSPs in the number of days to heading across environments was 1.5 days, with the landraces showing the longest cycle. This small difference in heading time reflects the criterion adopted to choose the accessions to be included in our panel, namely a similarity in phenology. The highest values for plant height and peduncle length were recorded in the 'ICARDA continental-dryland' accessions (GSP 5) and the landraces (GSP 8), while the lowest ones in the Italian and Spanish cultivars (GSPs 6 and 7).

3.2.2. Phenotypic structure

3.2.2.1. Based on multivariate data across environments. The dendrogram classifying accessions based on the mean phenotypic data of the six traits across environments is shown in Fig. 1. In order to derive a similar number of phenotypic and genetic subpopulations, and thus analogous degrees of freedom for the statistical tests, the accessions were grouped into nine subpopulations, as indicated by the vertical line in Fig. 1. The number of accessions included in each subpopulation ranged from 1 to 73, but most subpopulations contained between 11 and 20 accessions.

3.2.2.2. Based on univariate data in each of the nine environments. The Corsten and Denis algorithm applied to phenotypic univariate data grouped the accessions into 3–8 subpopulations depending on the trait (three for thousand kernel weight, four for plant height, six for days to heading and peduncle length and seven for number of grains per spike). This procedure also allowed the grouping of environments with similar effects on each trait. Fig. 2 shows the dendrogram of accessions (upper part) and environments (lower part) clustered according to yield data. The vertical line shows the clustering of accessions in eight subpopulations and environments in four groups, explaining ca. 42% of the GE interaction. Environments corresponding to experiments 8 (Tns-i05), 1 (Itl1-r04) and 9 (Tns-r05) clustered separately from each other and apart from the rest.

3.3. Comparison of genetic and phenotypic clustering

3.3.1. Association between matrices based on genetic and phenotypic data

The relationship between the distance matrices based on genetic and phenotypic multivariate data gave a value of 0.1554, with a very low probability of spurious association (Table 4). Similarly,

Table 3
Mean values ±SD of phenotypic traits for the eight subpopulations. Data are means across nine environments.

Subpopulation	Yield (t ha^{-1})	Grains spike^{-1} (no.)	Thousand kernel weight (g)	Days to heading (d)	Plant height (cm)	Peduncle length (cm)
1. CIMMYT-'Altar 84'	4.54 ± 1.91a	36.6 ± 9.2a	36.6 ± 6.8f	127.6 ± 23.1bcd	78.5 ± 13.3c	14.2 ± 4.6c
2. CIMMYT-'Yavaros 79'	4.52 ± 1.91a	34.3 ± 7.9cd	40.0 ± 7.7a	127.1 ± 22.9cd	77.2 ± 13.0d	13.7 ± 4.3cd
3. ICARDA-High yield	4.47 ± 1.90ab	34.6 ± 8.6bcd	39.6- ± 8.0ab	127.4 ± 23.2bcd	78.4 ± 13.4c	14.2 ± 4.5c
4. Moroccan + ICARDA – Temperate dryland	4.43 ± 1.81b	34.9 ± 8.5bc	39.0 ± 7.3bc	127.3 ± 23.0cd	76.7 ± 12.7d	13.4 ± 4.4de
5. ICARDA – Continental-dryland	4.49 ± 1.86ab	35.2 ± 8.8b	38.9 ± 6.8cd	127.0 ± 22.3d	86.3 ± 19.9a	16.9 ± 6.6a
6. Italian-'Valnova'	4.47 ± 1.86ab	34.3 ± 8.1cd	40.2 ± 7.7a	127.9 ± 22.1abc	75.6 ± 12.5e	13.1 ± 4.2ef
7. Spanish	4.41 ± 1.87b	34.8 ± 8.2bcd	38.1 ± 7.6e	128.1 ± 24.0ab	75.9 ± 12.9de	12.7 ± 4.1f
8. Landrace-derived	4.18 ± 1.75c	34.0 ± 7.6d	38.3 ± 7.2de	128.5 ± 22.9a	84.6 ± 20.1b	15.4 ± 6.8b

Means within columns with the same letters are not significantly different at $P < 0.05$ according to Duncan's test.

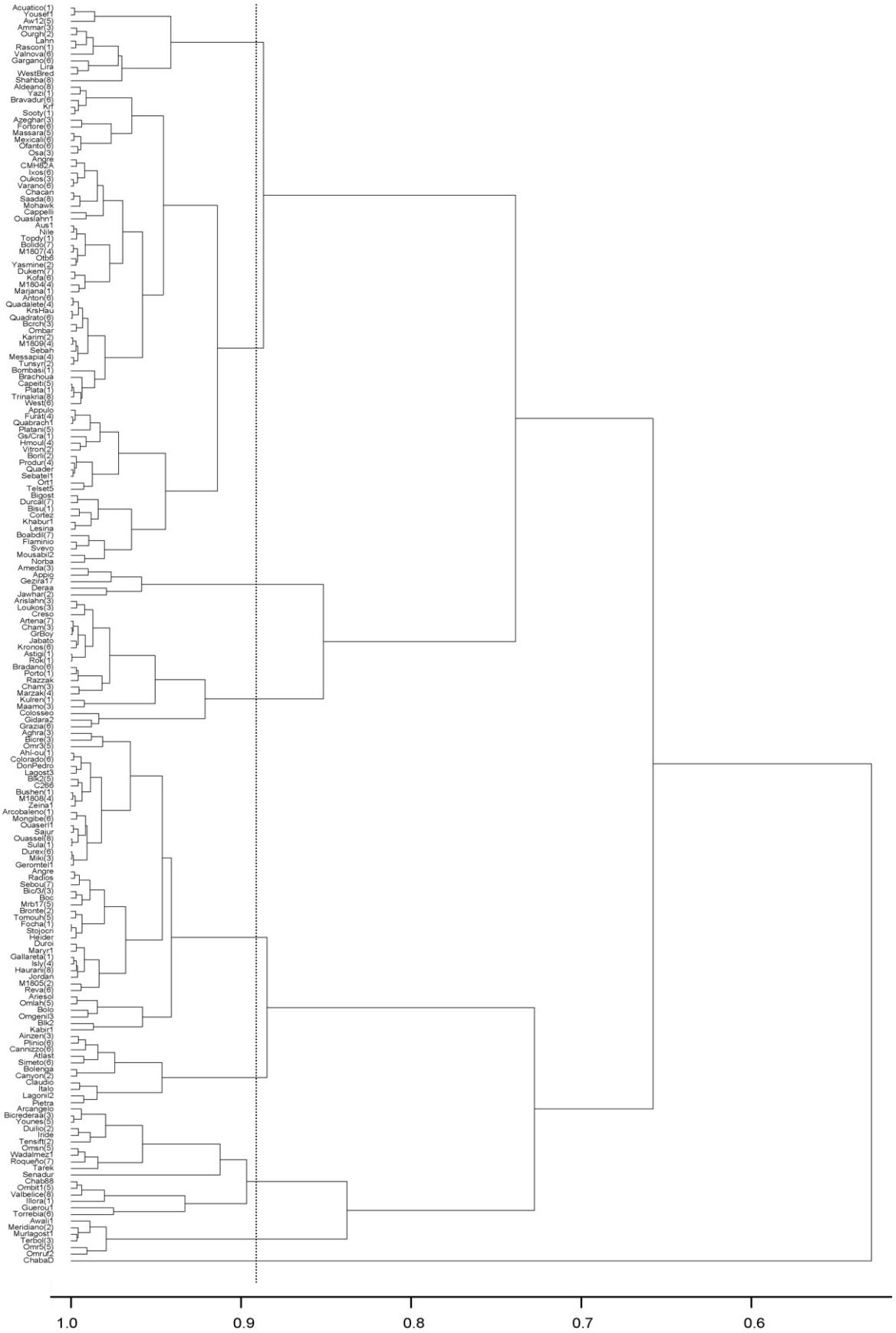

Fig. 1. Dendrogram showing the clustering of the 191 durum wheat accessions obtained by applying the furthest neighbor algorithm to the mean phenotypic data of six agronomic traits across nine environments. The vertical line identifies nine phenotypic subpopulations. Accessions followed by numbers between parentheses indicate the subpopulation estimated by the STRUCTURE software (Table 2a). Accessions without a number in parenthesis are those with less than 50% probability of belonging to any GSP (unstructured) according to Table 2b.

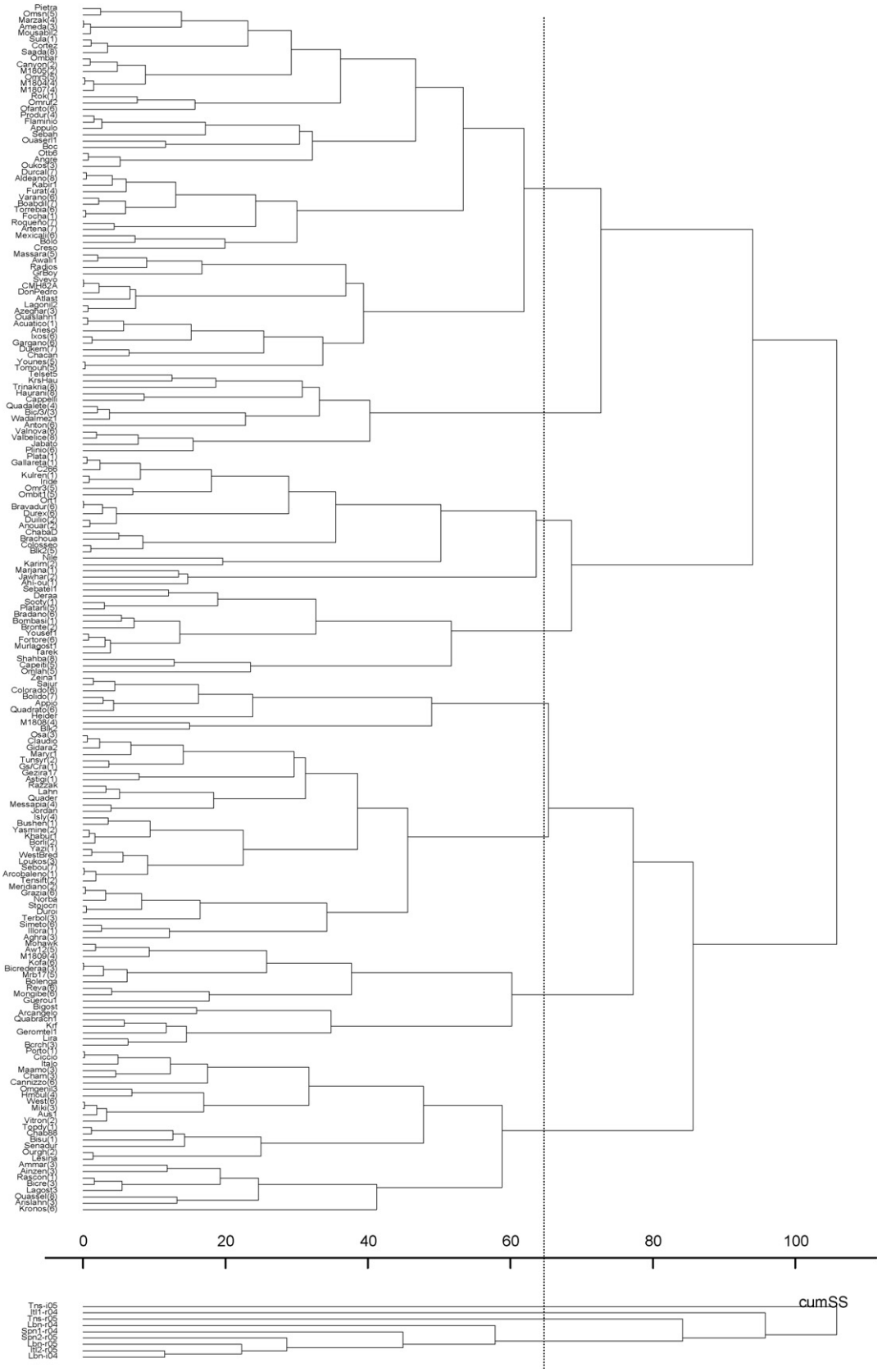

Fig. 2. Dendrograms of simultaneous clustered accessions (upper panel) and environments (lower panel) obtained by applying the Crosten and Dennis (1990) algorithm to the yield data of 191 durum wheat accessions in nine environments. Accessions followed by numbers between parentheses indicate the subpopulation estimated by the STRUCTURE software (Table 2a). Accessions without a number in parenthesis are those with less than 50% probability of belonging to any GSP (unstructured) according to Table 2b.The vertical bar identifies eight groups of accessions (phenotypic subpopulations) and four groups of environments: (i) environment 8 (Tns-i-05), (ii) environment 1 (Itl-1-r04), (iii) environment 9 (Tns-r-05), and (iv) all the remaining environments. The horizontal axis indicates the cumulative sum of squares of the GE interaction.

Table 4
Correlation analysis between the similarity matrix based on molecular data and the similarity matrices calculated from the six phenotypic variables (multivariate), and from each trait independently (univariate).

Matrices compared	Mantel test	
	Value	*P*-value[a]
Genetic–Phenotypic multivariate	0.1554	<0.001
Genetic–Phenotypic yield	0.0174	0.110
Genetic–Phenotypic number of grains per spike	0.0148	0.143
Genetic–Phenotypic thousand kernel weight	0.0263	0.0176
Genetic–Phenotypic days to heading	0.0262	0.0960
Genetic–Phenotypic plant height	0.1157	<0.001
Genetic–Phenotypic peduncle length	0.0647	0.008

[a] Estimated *P*-value based on 5000 permutations.

the relationships between the genetic distance matrix and the distance matrices for plant height, peduncle length and thousand kernel weight probably reflected causal associations. As expected, plant height and peduncle length were strongly correlated ($r = 0.82$, $P < 0.001$). In contrast, 11% of the permutations in the matrices based on genetic and yield data had associations greater or equal to those of the original matrices, suggesting that the association between them was probably spurious. The Mantel test also revealed that the associations between the SSR-based distance matrices and the matrices of distances based on grains per spike and days to heading were most likely spurious.

3.3.2. Analyses of variance

The environmental conditions exerted the most important effect on the six traits studied since they explained from 76.3 (for peduncle length) up to 98.5% (for days to heading) of the observed variability (Table 5). The genotype effect and the GE interaction were also significant in all cases. The percentage of variation explained by the genotype ranged from 0.37% for days to heading to 12.6% for plant height, and the percentage of variation explained by GE interaction ranged from 1.15 (for days to heading) to 12.5% (for peduncle length).

In order to assess which clustering method best explained the observed phenotypic variation, the sum of squares of the G and the GE effects were partitioned according to the structures defined by each of the clustering methods used. The results showed that the 113 genotypes assigned to one of the sub-populations (structured) explained from 51.0 to 72.2% of the sum of squares of the genotype effect for the different traits, whereas the 79 unassigned genotypes (unstructured) explained the remaining portion (from 27.8 to 49%; Table 5). Differences in yield between subpopulations explained from 22.9 to 30.3% of the sum of squares of the genotype effect depending on the clustering method, although most of the genetic variation was explained by variability within subpopulations. For the number of grains per spike, thousand kernel weight, days to heading and peduncle length, the phenotypic multivariate clustering method was the most efficient, explaining from 52.3 to 75.1% of the observed variability for these traits. The efficiency of the phenotypic univariate method varied according to the trait, being useful for plant height and peduncle length (explaining 72.8 and 68.8% of the sum of squares of the genotype effect, respectively), adequate for days to heading (55.3%), but not so relevant for grains per spike (43.5%), kernel weight (31.6%) and particularly for yield (28.3%).

The genetic structure of the population based on microsatellite data explained from 52.8 to 67.7% of the GE interaction for the traits analyzed (Table 5). However, the largest percentage of the environment × structured accession interaction was explained by the within subpopulation components. The clustering of the accessions based on phenotypic multivariate data generally offered a poor explanation of the GE interaction. Despite being highly significant, variability between subpopulations only explained from 7.3 (for yield) to 48.3% (for plant height) of the GE interaction. Differences between the subpopulations derived from univariate phenotypic data explained from 28.5 (for kernel weight) to 74.9% (for days to heading) of the GE interaction that was highly significant for all traits.

3.4. Genetic and phenotypic diversity

The clustering of accessions in the tree obtained using SSR-based dissimilarities (Fig. 3) gave a cophenetic index of 0.82, and largely confirmed the associations defined by the STRUCTURE software. It is notable that all accessions with a >86% probability of belonging to a GSP (see Table 2a) clustered in the same branch of the tree, but below this threshold the grouping of accessions did not follow the STRUCTURE results strictly. For example, accessions from GSP 1 clustered in three different branches (Fig. 3). Even though cv. 'Canyon', 'Borli', 'Tunsyr-1' and 'Tensift-1' belonged to GSP 2, they clustered separately from the remaining members of this group. Fourteen of the 17 accessions included in GSP 3 clustered very tightly, whereas cultivars 'Arislahn-5', 'Azeghar' and 'Osa' were closer to unclassified accessions. In GSP 4 and GSP 5, respectively, 73 and 85% of the accessions clustered together. However, cultivars 'Capeiti-8' and 'Platani' were placed in the same branch as 'Appulo', 'Ciccio' and 'Capelli', and the two Italian accessions 'Mongibello' and 'Cannizzo'. The group of 23 accessions genetically close to the Italian founder 'Valnova' was very compact with only three outliers. The Spanish cultivars 'Boabdil', 'Bolido' and 'Durcal' clustered close to the Italian variety 'Svevo', which was not included in the genetic structure. Similarly 'Artena', 'Roqueño' and 'Sebou', the latter from ICARDA, clustered together and were placed close to cultivar 'Appio'. All the landrace-derived accessions, with the exception of 'Ouassel', clustered very tightly (Fig. 3).

Most of the accessions included in GSPs 1, 5 and 8 clustered together in the tree built from phenotypic data (Fig. 4), which gave a cophenetic index of 0.83. Thus, 17 of the 21 GSP 1 accessions clustered together according to their phenotypic performance. However, they were also close to accessions belonging to other GSPs, and even to some genetically unstructured accessions. Similarly, 10 of the 13 GSP 5 accessions, and 5 of the 7 GSP 8 accessions clustered together, but were placed close to accessions (e.g. 'Anton' and 'Grazia') from GSP 6 (Fig. 4). There were also several cases of accessions belonging to the same GSP clustering together in the tree built from phenotypic data (e.g. cultivars 'Tunsyr-1', 'Duilio', 'Karim' and 'Ourg', from GSP 2; 'Durex' and 'Mexicali' from GSP6; 'Loukos-1' and 'Maamouri-1' from GSP 3; 'Artena' and 'Roqueño' from GSP 7; 'Furat-1' and 'Quadalete' from GSP 4). However, in most cases, the accessions included in the same GSP were distributed along the different branches of the tree; the same was true for the unclassified accessions.

4. Discussion

4.1. Environmental

Yield was not associated to water input nor before or after heading. The only environmental variable that entered in the linear regression model, showing a negative relationship to yield, was the maximum temperature from emergence to heading, which explained a very high percentage (59%) of yield variation. The maximum temperature from emergence to heading was also a critical factor to identify groups of environments that maximally explained the yield GE interaction, since in the dendrogram obtained by applying the Corsten and Denis algorithm to the yield obtained by each accession on each environment, the three environments with the lowest maximum temperatures from emergence to heading (1, 8 and 9) clustered separately one from the other and from the

Table 5
Analysis of variance for the six traits for 191 durum wheat accessions grown in nine environments across the Mediterranean Basin. The genotype effect and the genotype × environment interaction are partitioned according to the three clustering methods used to group the accessions in subpopulations: (i) clustering based on molecular data (genetic structure), (ii) clustering based on phenotypic data of the six traits across environments (phenotypic structure multivariate), and (iii) clustering based on phenotypic data of each trait on each environment (Corsten and Denis algorithm partitioning, phenotypic structure univariate). GSP = Genetic subpopulations, PSP = Phenotypic subpopulations.

Source of variation	Yield (t ha^{-1})				NGS				TKW (g)				Days to heading				Plant height (cm)				Peduncle length (cm)			
	d.f.	SS	%SS	−log(*P*)	d.f.	SS	%SS	−log(*P*)	d.f.	SS	%SS	−log(*P*)	d.f.	SS	%SS	−log(*P*)	d.f.	SS	%SS	−log(*P*)	d.f.	SS	%SS	−log(*P*)
Environment	8	5846	97.7	>100	8	105843	89.3	>100	8	85172	83.7	>100	8	876942	98.5	>100	8	287733	78.5	>100	8	30373	76.3	>100
Genotype	190	29.3	0.49	15.6	190	2558	2.16	12.2	190	7935	7.80	118	190	3324	0.37	23.0	190	46095	12.6	185	190	4470	11.2	115
Genetic structure																								
Structured genotypes	112	19.0	64.9	12.2	112	1770	69.2	8.69	112	4739	59.7	75.2	112	1695	51.0	13.0	112	26887	58.3	112	112	3228	72.2	73.7
Between GSP	7	6.71	35.3	7.20	7	732	41.4	9.29	7	1670	35.2	7.19	7	164	9.69	0.85	7	12168	45.3	10.8	7	1467	45.4	10.8
Within GSP	105	12.3	64.7	4.46	105	1037	58.6	2.81	105	3069	64.8	48.3	105	1531	90.3	9.38	105	14719	54.7	66.6	105	1761	54.6	47.9
Unstructured genotypes	78	10.3	35.1	4.65	78	788	30.8	4.22	78	3196	40.3	45.0	78	1629	49.0	11.0	78	19207	41.7	74.4	78	1242	27.8	40.2
Phenotypic structure multivariate																								
Between PSP	8	8.89	30.3	10.7	8	1337	52.3	24.9	8	5956	75.1	50.1	8	2450	73.7	48.1	8	28821	62.5	34.3	8	3079	68.9	41.5
Within PSP	182	20.4	69.7	5.67	182	1221	47.7	0.33	182	1979	24.9	10.2	182	874	26.3	0.00	182	17273	37.5	58.0	182	1390	31.1	17.2
Phenotypic structure univariate																								
Between PSP	7	8.30	28.3	10.2	6	1113	43.5	19.9	2	2509	31.6	15.5	5	1840	55.3	29.9	3	33538	72.8	51.8	5	3074	68.8	44.2
Within PSP	183	21.0	71.7	6.20	184	1445	56.5	1.21	188	5426	68.4	73.3	185	1484	44.7	1.34	187	12557	27.2	32.8	185	1396	31.2	16.9
Genotype × Environment	1520	105.8	1.77	#	1520	10136	8.55	#	1520	8616	8.47	#	1520	10198	1.15	#	1520	32731	8.93	#	1520	4974	12.5	#
Genetic structure																								
Environment × Structured	896	62.1	58.7	NT	896	6670	65.8	NT	896	4852	56.3	NT	896	5385	52.8	NT	896	18611	56.9	NT	896	3366	67.7	NT
Env × Between GSP	56	8.44	13.6	6.64	56	972	14.6	7.94	56	617	12.7	5.57	56	827	15.3	9.01	56	8333	44.8	74.5	56	1040	30.9	38.0
Env × Within GSP	840	53.7	86.4	NT	840	5698	85.4	NT	840	4235	87.3	NT	840	4559	84.7	NT	840	10279	55.2	NT	840	2326	69.1	NT
Env × Unestructured	624	43.7	41.3	NT	624	3467	34.2	NT	624	3764	43.7	NT	624	4812	47.2	NT	624	14119	43.1	NT	624	1607	32.3	NT
Phenotypic structure multivariate																								
Env × Between PSP	64	7.72	7.30	3.78	64	1340	13.2	16.8	64	1744	20.2	37.8	64	3160	31.0	77.9	64	15811	48.3	163	64	1371	27.6	64.1
Env × Within PSP	1456	98.1	92.7	NT	1456	8796	86.8	NT	1456	6872	79.8	NT	1456	7038	69.0	NT	1456	16919	51.7	NT	1456	3603	72.4	NT
Phenotypic structure univariate																								
Env × Between PSP	56	45.0	42.6	137	48	4717	46.5	164	16	2457	28.5	97.0	40	7639	74.9	>100	24	19480	59.5	272	40	2336	47.0	173
E × Within PSP	1464	60.7	57.4	NT	1472	5419	53.5	NT	1504	6159	71.5	NT	1480	2559	25.1	NT	1496	13251	40.5	NT	1480	2638	53.0	NT
Total	1718	5981	1718	118538	1718	101723	1718	890463	1718	366558	1718	39817												

'#' Significant at $P < 0.001$ using the median intrablock error (shown in Table 1 for yield). NT, not testable.

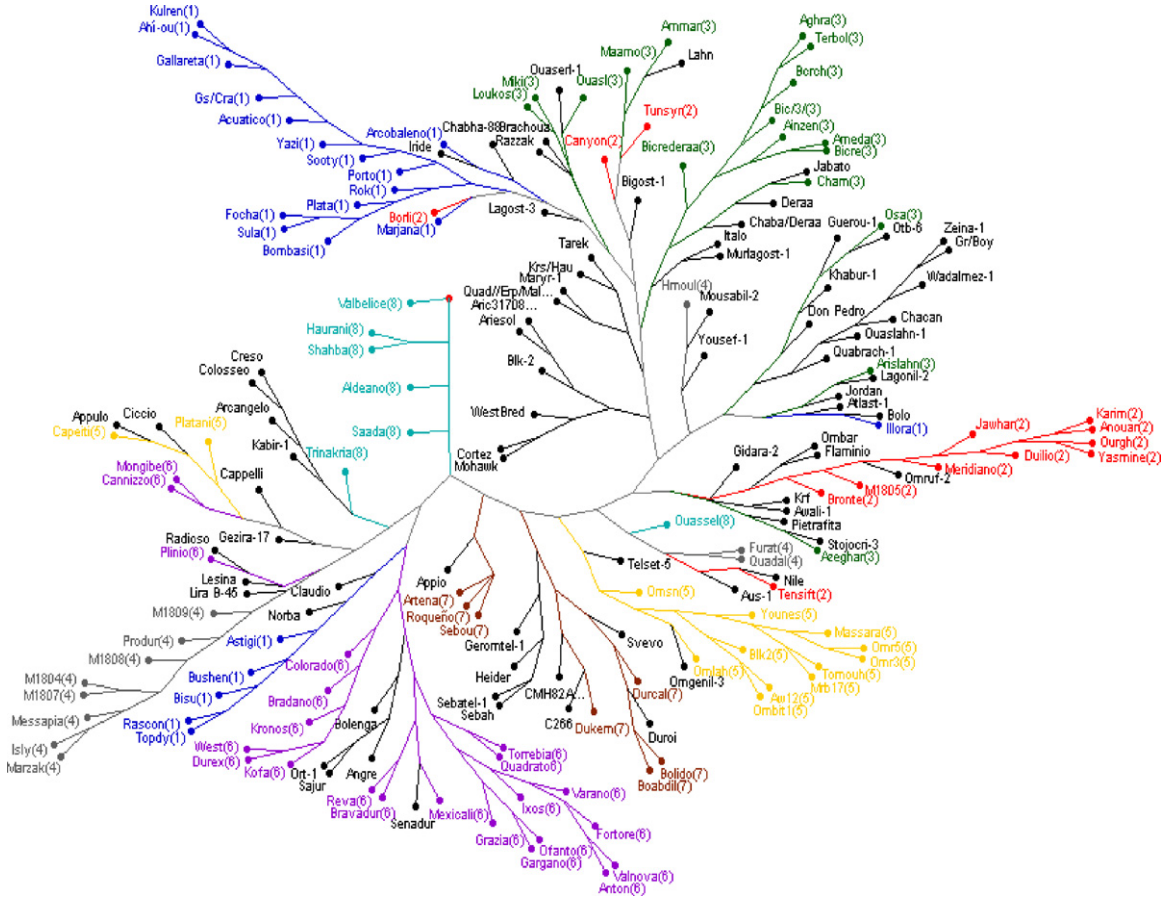

Fig. 3. Un-rooted hierarchical clustering tree based on dissimilarities between the accessions from 72 SSR data. Numbers between parentheses indicate the GSP estimated by the STRUCTURE software. Accessions without a number in parenthesis are those with less than 50% probability of belonging to any GSP (unstructured) according to Tables 2a and 2b.

rest. The large negative impact of maximum temperature before heading on yield may reflect its effect on shortening the growth cycle, thus reducing the time available for accumulating biomass and resources for grain filling (Mitchell et al., 1993), hence reducing the yield potential (Amir and Sinclair, 1991).

4.2. Genetic structure and its breeding meaning

The eight GSPs resulting from the SSR marker analysis are useful to depict the genetic structure and diversity of the durum germplasm currently grown in the Mediterranean Basin. They also provide a picture of the genetic relationships between the germplasm used in different countries within the region.

CIMMYT germplasm was clustered in two subpopulations corresponding to two generations of durum wheat varieties. 'Yavaros 79', which showed the largest kernel weight, was the result of breeding efforts to improve the agronomic components associated with high yield potential and wide adaptation (Royo et al., 2009). The release, with different names, of a number of 'Yavaros 79' sister ('s') lines in different countries (Karim' in Tunisia, 'Vitron' in Spain, 'Yasmine' in Morocco), and others closely related to it such as 'Duilio' (which shares with 'Yavaros' the parent Anhinga/Flamingo) in Italy, confirms its wide adaptation and the large impact of durum CIMMYT germplasm in the Mediterranean Basin. The high grain weight of 'Yavaros 79' may be a key contributor to its reported yield stability (Pfeiffer et al., 2000) since it has been shown that durum wheat cultivars characterized by high grain weight (Royo et al., 2007) are more stable across environments (Royo et al., 2008). In contrast, 'Altar 84' belongs to a new generation of durum wheat varieties characterized by a balanced increase in all yield components (Royo et al., 2009). The high yield potential of the Altar 84-derived germplasm reflects its large sink size, which agrees with the well-known association between the number of grains per spike and the increases in the yield potential of durum wheat during the 20th century (Giunta et al., 2007; Royo et al., 2007). ICARDA germplasm clustered in three subpopulations, one of them closely related to Moroccan accessions, which illustrates the involvement of ICARDA lines adapted to temperate dry land areas in the development of Moroccan germplasm. The ICARDA lines adapted to continental-dryland areas were characterized by tall plants with very long peduncles, keeping with the reported positive association between peduncle length and grain yield under drought-stressed Mediterranean conditions (Nachit and Elouafi, 2004).

The Italian durum wheat genetic pool is considered one of the most, if not the most, representative within the Mediterranean Basin, due to the early and continued efforts devoted to durum wheat breeding in Italy since the beginning of the 20th century (Royo et al., 2009). The clustering of the desert durum-USA accessions within the Italian genetic subpopulation reflects the introduction of Italian germplasm in North American breeding programs, as a way to widen the genetic background and to improve grain quality (Royo et al., 2009). Some Spanish accessions were closely related to the CIMMYT subpopulations, revealing the significance of CIMMYT hallmark founders in the release of Spanish varieties. However, a small group of Spanish accessions formed a specific subpopulation, suggesting the presence of alter-

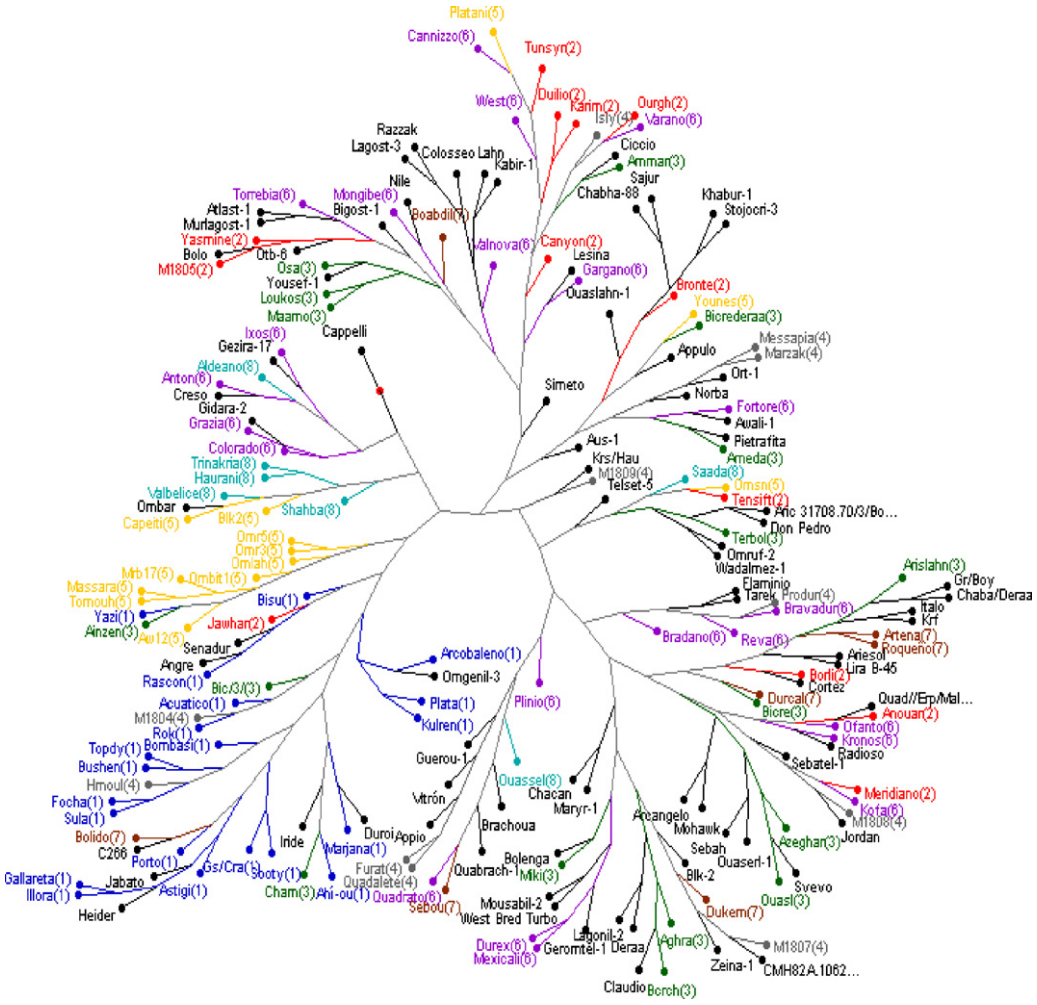

Fig. 4. Un-rooted hierarchical clustering tree based on Euclidean distances of the standardized mean phenotypic data across environments. Numbers between parentheses indicate the subpopulation estimated by the STRUCTURE software. Accessions without a number in parenthesis are those with less than 50% probability of belonging to any GSP (unstructured) according to Tables 2a and 2b.

native sources of germplasm in the development of some cultivars. Landrace-derived genotypes made up a specific subpopulation characterized by low productivity, tall plants and low number of grains per spike, most likely associated to the absence of *Rht-B1* dwarfing gene, as common in landraces and old varieties (Royo et al., 2007).

4.3. Relationship between genetic and phenotypic structures

Correlation analysis between the matrices of distances obtained using SSR and phenotypic data revealed some significant, but weak, associations between genetic and phenotypic structures. Significant relationships were obtained between the genetic structure of the population and the clustering of the accessions based on the data across environments of the six phenotypic traits, but the predictive value of the genotypic-multivariate phenotypic-similarity was only 2.4%. Significant associations were also found between the genetic structure and the phenotypic structures based on plant height, peduncle length and thousand kernel weight, but their predictive values were even lower than that of the multivariate structure. No association was found between the genetic structure and the phenotypic structures for yield, grain per spike and days to heading. These results suggest a very slow possibility to predict the phenotypic performance from random molecular markers.

4.4. Utility of genetic and phenotypic structures in explaining the agronomic performance

The ANOVA showed that a larger percentage (51.0–72.2%) of the genotypic variation for the investigated traits was explained by the 113 genetically structured accessions as compared to the 79 unstructured ones. However, the percentage of genetic variability explained by differences between GSPs was generally low, ≤33% for all the six phenotypic traits, and in all cases lower than that explained by the variability existing within GSPs, which suggests that the genetic clustering did not accurately reflect the phenotypic performance of the entire population. Only a limited portion of the environment × structured genotype interaction was ascribable to the between GSP component, thus suggesting the inadequacy of genetic clustering also to explain the GE interaction.

The clustering accessions on the basis of phenotypic multivariate data was, as compared to the molecular classification, a much better procedure to explain the genotypic effects since in the first case the percentage of the sum of squares between phenotypic subpopulations was higher than that between genetic subpopulations for all the traits (e.g. 30.3 vs. 22.9% for yield and 75.1 vs. 21.0% for kernel weight). The percentage of genetic variability explained by differences between phenotypic subpopulations ranged from 30.3 to 75.1% for the multivariate structure, and from 28.3 to 72.8% for the univariate structure. Nevertheless, the classification of acces-

sions based on phenotypic multivariate data was better than that based on univariate data when explaining genotypic differences in kernel weight, days to heading and number of grains per spike, while it was similar for peduncle length and yield but less robust for plant height. These results indicate that none of the clustering methods we used was superior for the complete set of traits but, in general, the procedure based on multivariate phenotypic data was the most useful clustering method among the three that were tested to explain genotypic differences. Moreover, none of the clustering methods we used was appropriate for yield; in fact, in all three cases differences between subpopulations did not explain more than 30% of yield variation among genotypes. This was probably due to the complex nature of the trait, which is also highly affected by environmental conditions (Jackson et al., 1996).

The phenotypic univariate clustering of accessions was consistently the most helpful in explaining GE interaction since it revealed from 28.5 to 74.9% of the GE interaction whereas the multivariate structure revealed just from 7.3 to 48.3%. These results may reflect the fact that multivariate phenotypic clustering was based on data from each genotype across environments, thus considering only the genotype effect, whereas univariate phenotypic clustering was based on data from each accession in each environment, hence taking into account genotypic and environmental effects.

Plant height and peduncle length performed quite differently from the other traits. In fact, the comparison of matrices obtained with genetic and phenotypic univariate data revealed that the grouping of accessions based on plant height and peduncle length were significantly related to the genetic structure of the population. Moreover, the phenotypic univariate structure explained a higher portion of the genotypic differences in plant height and peduncle length, as compared to the other traits. The differences in the performance of these two traits may be consequence of their high heritability (Collaku, 1994), which has been associated to the presence of the *Rht-B1b* dwarfing gene (Borner et al., 1997). Actually, a major QTL, mapping directly to the *Rht-B1* locus on chromosome arm 4BS, has been reported to account for up to 49% of the genotypic variance in peduncle length and plant height (Rebetzke et al., 2001). This agrees with the relatively low environmental impact and the larger effect of genotypic variation on plant height and peduncle length compared to the other traits, as revealed by the ANOVA.

4.5. Genetic and phenotypic diversity

The classification of the accessions obtained using the clustering method based on SSR dissimilarities (implemented in DARWin) was in agreement with the population structure results (obtained with the model-based cluster analysis implemented in STRUCTURE) only when considering accessions with a high probability (> 86%) of belonging to a specific subpopulation. However, below this threshold, the agreement was not so good; as an example accessions genetically close to 'Altar 84' and included in the same genetic subpopulation (GSP1) were placed by DARWin into two different branches. The Italian cultivars 'Iride' (pedigree Altar84/Ares), 'Norba' and 'Claudio' were also placed close to GSP1, suggesting the utilization by Italian breeders of 'Altar 84' or its derivatives in their crossing schemes as a strategy to increase yield potential. Another case of discrepancy between the genetic structure and the diversity analysis based on SSR data occurred in GSP 5; cultivars 'Capeiti-8' (derived from the cross Capelli/Eiti and released in 1940) and 'Platani' (derived from the cross Valnova/Capeiti and released in 1995) clustered separately from the other members of GSP 5 in the tree formed from genetic dissimilarities, while being close to the unstructured cultivars 'Apullo' (Cappelli/Grifoni/Capeiti-8, released in Italy in 1973) and 'Ciccio' (F6 Appulo/Valnova//Valforte/Patrizio, released in Italy in 1996). The fact that these four accessions share the old ancestor 'Capeiti-8' (Rascio et al., 1992), probably used as donor for drought-tolerance, substantiates their position in a common branch. Nevertheless, the different number of SSRs used in calculations by STRUCTURE and DARWin may also explain these discrepancies.

In the tree based on SSR dissimilarities, the ICARDA germplasm specifically adapted to high-yielding areas (GSP 3) formed, with only few exceptions, a very compact group that was genetically close to GSP 1 (including Altar84-related accessions also adapted to areas of high yield potential) and to some ICARDA accessions not included in any of the eight GSPs. In this respect, it is noticeable that breeders from CIMMYT and ICARDA extensively used also the lines 'Ruffo', 'Flamingo' and 'Mexicali75' as parents in their breeding programs; thus one or more of these lines may be common ancestors of the accessions grouped in this branch.

In the tree obtained using molecular data, 7 of the 10 accessions from irrigated areas of Arizona and California clustered with the majority of the Valnova-related Italian accessions included in GSP 6 and with some CIMMYT lines. These results together with the fact that cv. 'Valnova' includes in its pedigree old CIMMYT germplasm suggest that similar CIMMYT germplasm could have been exploited in Italian and American breeding programs. The knowledge of the genetic proximity between accessions gives clues about the heterosis that can be expected when making crosses among them in order to create polymorphic populations, given at the same time, information on the likely origin of the accessions.

The tree formed from Euclidean distances of the standardized phenotypic means across environments showed that genetic relatedness only rarely matched analogous agronomic performance. GSP 1 accessions (genetically close to cv. 'Altar 84') were the most homogeneous in terms of agronomic performance, given that 17 of the 21 accessions clustered in the same branch of the phenotypic tree. The ICARDA accessions included in GSP 5 (germplasm for continental-dryland areas) comprised another agronomically compact group since 10 out of 13 accessions clustered in the same branch. GSP 8, predominantly including landrace-derived accessions, was also agronomically homogeneous. The information given by this tree, which bunches accessions with similar overall agronomic performance, may be valuable to ascertain distances between given accessions and varieties very well known as being largely cultivated, which may help in planning crosses on breeding programs.

5. Concluding remarks

Our set of random SSR markers was useful to assign many accessions to subpopulations and describe the genetic structure of the collection. The genetic relatedness between CIMMYT lines and several varieties successful in the Mediterranean Basin illustrates the important contribution of CIMMYT germplasm in the region. Even so, the two subpopulations closely related to CIMMYT germplasm showed contrasting yield formation strategies. Despite the high grain yield reached by both subpopulations, in the subpopulation containing accessions related to 'Yavaros 79' the high yielding level was mainly achieved through grain weight, while in the subpopulation close to 'Altar 84' it was mainly due to a large number of grains per spike, which boosted the sink potential.

A significant relationship was detected between the genetic similarities among accessions and their phenotypic resemblances in terms of the agronomic traits. However, this association was not strong enough to properly explain the observed phenotypic variability, probably because the SSR markers were randomly chosen and their association with the studied traits was unknown. Traits regulated by major genes, such as plant height, were the only ones that led to a phenotypic structure that was significantly associated with the genetic one.

None of the clustering methods we used explained more than 30% of yield variations due to the genotype effect. However, the structure of the population built from the mean data of the six phenotypic traits across environments explained a large portion of the variability due to the genotype effect for most of the traits, thus showing to be much more effective than the molecular classification to explain the variability due to genotypes. The genetic structure of the population was also inappropriate to explain the GE interaction, which was consistently best explained by the structures based on univariate phenotypic data covering the whole set of environments.

Acknowledgements

The financial contribution of the European Union (IDuWUE project: INCO contract number ICA3-CT-2002-10028) is gratefully acknowledged. The authors would like to thank Drs Ignacio Romagosa and Fred A. van Eeuwijk for their assistance with statistical analysis. The Centre UdL-IRTA forms part of the Centre CONSOLIDER INGENIO 2010 on Agrigenomics funded by the Spanish Ministry of Education and Science.

References

Allen, R.G., Pereira, L.S., Raes, D., Smith, M., 1998. Crop evapotranspiration: guidelines for computing crop requirements. FAO Irrigation and Drainage Paper No. 56. FAO, Rome, Italy.

Amir, J., Sinclair, T.R., 1991. A model of the temperature and solar-radiation effects on spring wheat growth and yield. Field Crops Res. 28, 47–58.

Annicchiarico, P., Royo, C., Bellah, F., Moragues, M., 2009. Relationships among adaptation patterns, morphophysiological traits and molecular markers in durum wheat. Plant Breed. 128, 164–171.

Araus, J.L., Bort, C., Steduto, P., Villegas, D., Royo, C., 2003. Breeding cereals for Mediterranean conditions: ecophysiological clues for biotechnology application. Ann. Appl. Biol. 142, 129–141.

Autrique, E., Nachit, M.M., Monneveux, P., Tanksley, S.D., Sorrells, M.E., 1996. Genetic diversity in durum wheat based on RFLPs, morphophysiological traits, and coefficient of parentage. Crop Sci. 36, 735–742.

Boggini, G., Palumbo, M., Calcagno, F., 1990. Characterization and utilization of Sicilian landraces of durum wheat in breeding programs. In: Srivastava, J.P., Damania, A.B. (Eds.), Wheat Genetic Resources: Meeting Diverse Needs. John Wiley & Sons, Chichester, UK, pp. 223–234.

Borner, A., Roder, M., Korzun, V., 1997. Comparative molecular mapping of GA insensitive Rht loci on chromosomes 4B and 4D of common wheat (*Triticum aestivum* L.). Theor. Appl. Genet. 95, 1133–1137.

Campbell, G.S., Mulla, D.J., 1990. Measurement of soil water content and potential. In: Stewart, B.A., Nielsen, D.R. (Eds.), Irrigation of Agricultural Crops. ASA-CSSA-SSSA, Madison, pp. 127–141.

Collaku, A., 1994. Selection for yield and its components in a winter-wheat population under different environmental-conditions in Albania. Plant Breed. 112, 40–46.

Corsten, L.C.A., Denis, J.B., 1990. Structuring interaction in two-way tables by clustering. Biometrics 46, 207–215.

Daaloul, A., Harrabi, M., Ammar, K., Abdennadher, M., 1990. Evaluation of durum wheat lines for yield, drought tolerance and Septoria resistance in Tunisia. In: Srivastava, J.P., Damania, A.B. (Eds.), Wheat Genetic Resources: Meeting Diverse Needs. John Wiley & Sons, Chichester, UK, pp. 187–194.

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

Donini, P., Law, J.R., Koebner, R.M.D., Reeves, J.C., Cooke, R.J., 2000. Temporal trends in the diversity of UK wheat. Theor. Appl. Genet. 100, 912–917.

Falush, D., Stephens, M., Pritchard, J.K., 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164, 1567–1587.

Federer, W.T., Reynolds, M., Crossa, J., 2001. Combining results from augmented designs over sites. Agron. J. 93, 389–395.

Fufa, H., Baenziger, P.S., Beecher, B.S., Dweikat, I., Graybosch, R.A., Eskridge, K.M., 2005. Comparison of phenotypic and molecular marker-based classifications of hard red winter wheat cultivars. Euphytica 145, 133–146.

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.

Jackson, P., Robertson, M., Cooper, M., Hammer, G., 1996. The role of physiological understanding in plant breeding, from a breeding perspective. Field Crops Res. 49, 11–37.

Jain, S.K., Qualset, C.O., Bhatt, G.M., Wu, K.K., 1975. Geographical patterns of phenotypic diversity in a world collection of durum wheats. Crop Sci. 15, 700–704.

Little, R.C., Milliken, G.A., Stroup, W.W., Wollfinger, R.D., 1996. SAS System for Mixed Models. SAS Institute Inc., Cary, North Carolina.

Mac Key, J., 2005. Wheat, its concept, evolution and taxonomy. In: Royo, C., Nachit, M.M., Di Fonzo, N., Araus, J.L., Pfeiffer, W.H., Slafer, G.A. (Eds.), Durum Wheat Breeding: Current Approaches and Future Strategies. Food Products Press, New York, pp. 3–61.

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.

Maccaferri, M., Sanguineti, M.C., Noli, E., Tuberosa, R., 2005. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. Mol. Breed. 15, 271–289.

Maccaferri, M., Sanguineti, M.C., Natoli, V., Araus, J.L., Ben Salem, M., Bort, J., Chenenaoui, C., De Ambrogio, E., García del Moral, L.F., Demontis, A., El-Ahmed, A., Maalouf, F., Machlab, H., Moragues, M., Motawaj, J., Nachit, M., Nserallah, N., Ouabbou, H., Royo, C., Tuberosa, R., 2006. A panel of elite accessions of durum wheat (*Triticum durum* Desf.) suitable for association mapping studies. Plant Gen. Res. 4, 79–85.

Maccaferri, M., Mantovani, P., Tuberosa, R., DeAmbrogio, E., Giuliani, S., Demontis, A., Massi, A., Sanguineti, M.C., 2008. A major QTL for durable leaf rust resistance widely exploited in durum wheat breeding programs maps on the distal region of chromosome arm 7BL. Theor. Appl. Genet. 117, 1225–1240.

Mantel, N., 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27, 209–220.

Mantovani, P., Maccaferri, M., Sanguineti, M.C., Tuberosa, R., Catizone, I., Wenzl, P., Thomson, B., Carling, J., Huttner, E., DeAmbrogio, E., Kilian, A., 2008. An integrated DArT-SSR linkage map of durum wheat. Mol. Breed. 22, 629–648.

Mitchell, R.A.C., Mitchell, V.J., Driscoll, S.P., Franklin, J., Lawlor, D.W., 1993. Effects of increased CO_2 concentration and temperature on growth and yield of winter-wheat at 2 levels of nitrogen application. Plant Cell Environ. 16, 521–529.

Moragues, M., Moralejo, M.A., Sorrells, M.E., Royo, C., 2007. Dispersal of durum wheat landraces across the Mediterranean basin assessed by AFLPs and microsatellites. Gen. Res. Crop Evol. 54, 1133–1144.

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

Nachit, M.M., Elouafi, I., 2004. Durum wheat adaptation in the Mediterranean dryland: breeding stress physiology and molecular markers. In: Rao, S.C., Ryan, J. (Eds.), Challenges and Strategies for Dryland Agriculture. CSSA Special Publication no. 32, pp. 203–218.

Peccetti, L., Annicchiarico, P., 1993. Grain yield and quality of durum wheat landraces in a dry Mediterranean region of Northern Syria. Plant Breed. 110, 243–249.

Perrier, X., Flori, A., Bonnot, F., 2003. Methods of data analysis. In: Hamon, P., Seguin, M., Perrier, X., Glazmann, J.C. (Eds.), Genetic Diversity of Cultivated Tropical Plants. CIRAD, Montpellier (France), pp. 31–63.

Pfeiffer, W.H., Sayre, K.D., Reynolds, M.P., 2000. Enhancing genetic gain yield potential and yield stability in durum wheat. In: Royo, C., Nachit, M.M., Di Fonzo, N., Araus, J.L. (Eds.), Durum Wheat Improvement in the Mediterranean Region: New Challenges. Options Méditerranéennes, Series A. 40, pp. 88–93.

Porceddu, E., 1976. Variation for agronomical traits in a world collection of durum wheat. Z. Pflanzenzüchtg 77, 314–327.

Pritchard, J., Stephens, M., Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics 155, 945–959.

Rascio, A., Platani, C., Di Fonzo, N., Wittmer, G., 1992. Bound water in durum wheat under drought stress. Plant Physiol. 98, 908–912.

Rebetzke, G.J., Appels, R., Morrison, A.D., Richards, R.A., McDonald, G., Ellis, M.H., Spielmeyer, W., Bonnett, D.G., 2001. Quantitative trait loci on chromosome 4B for coleoptiles length and early vigour in wheat (*Triticum aestivum* L.). Aust. J. Agric. Res. 52, 1221–1234.

Romagosa, I., van Eeuwijk, F.A., Thomas, W.T.B., 2009. Statistical analyses of genotype by environment data. In: Carena, M.J. (Ed.), Handbook of Plant Breeding, vol. 3. Springer Science + Business Media, Cereals, pp. 291–331.

Royo, C., Alvaro, F., Martos, V., Ramdani, A., Isidro, J., Villegas, D., García del Moral, L.F., 2007. Genetic changes in durum wheat yield components and associated traits in Italian and Spanish varieties during the 20th century. Euphytica 155, 259–270.

Royo, C., Elias, E.E., Manthey, F.M., 2009. Durum wheat breeding. In: Carena, M.J. (Ed.), Handbook of Plant Breeding, vol. 3. Springer Science + Business Media, Cereals, pp. 199–226.

Royo, C., Martos, V., Ramdani, A., Villegas, D., Rharrabti, Y., García del Moral, L.F., 2008. Changes in yield and carbon isotope discrimination of Italian and Spanish durum wheat during the 20th century. Agron. J. 100, 352–360.

Saghai Maroof, M.A., Solima, K.M., Jorgenson, R.A., Allard, R.W., 1984. Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics. PNAS 81, 8014–8018.

SAS Institute Inc., 2000. SAS Online Doc (R) Ver 8.

Soleimani, V.D., Baum, B.R., Johnson, D.A., 2002. AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat [*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.]. Theor. Appl. Genet. 104, 350–357.

Somers, D.J., Banks, T., DePauw, R., Fox, S., Clarke, J., Pozniak, C., McCartney, C., 2007. Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. Genome 50, 557–567.

Song, Q.J., Shi, J.R., Singh, S., Fickus, E.W., Costa, J.M., Lewis, J., Gill, B.S., Ward, R., Cregan, P.B., 2005. Development and mapping of microsatellite (SSR) markers in wheat. Theor. Appl. Genet. 110, 550–560.

Srivastava, J.P., 1987. Barley and wheat improvement 1 for moisture-limiting areas in West Asia and North Africa. In: Srivastava, J.P., Porceddu, E., Acevedo, E., Varma,

S. (Eds.), Drought Tolerance in Winter Cereals. John Wiley & Sons, Chichester, UK, pp. 65–78.
van Hintum, T.J.L., Elings, A., 1991. Assessment of glutenin and phenotypic diversity of Syrian durum wheat landraces in relation to their geographical origin. Euphytica 55, 209–215.
van Hintum, T.J.L., Haalman, D., 1994. Pedigree analysis for composing a core collection of modern cultivars, with examples from barley (*Hordeum vulgare*). Theor. Appl. Genet. 88, 70–74.
Zadoks, J.C., Chang, T.T., Konzak, C.F., 1974. A decimal code for the growth stage of cereals. Weed Res. 14, 415–421.