From IWIS-Bib

TrDR IT Colosseo **Cultivar Source: ?, ? IT** **CID:? SID:?**

Marco Maccaferri, Maria Corinna Sanguineti, Enrico Noli, Roberto Tuberosa. Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Molecular Breeding* 15 (3): 271-289 (2005) See Table 1 (pages 274-276) for pedigree and year of registration.

Molecular Breeding (2005) 15: 271–289
DOI 10.1007/s11032-004-7012-z

Population structure and long-range linkage disequilibrium in a durum wheat elite collection

Marco Maccaferri, Maria Corinna Sanguineti*, Enrico Noli and Roberto Tuberosa
*Department of Agroenvironmental Science and Technology, University of Bologna, Viale Fanin n. 44, Bologna 40127, Italy; *Author for correspondence (e-mail: maria.sanguineti@unibo.it)*

Received 11 June 2004; accepted in revised form 1 December 2004

Key words: Durum wheat, Genetic distance, Germplasm, Linkage disequilibrium, SSR, *Triticum durum* Desf

Abstract

A collection of 134 durum wheat accessions, mainly including cultivars (cvs.) representative of the major gene pools, was assembled and characterized with 70 SSRs for genetic diversity and level of long-range linkage disequilibrium (LD). Results of both a distance-based and a model-based (Bayesian) cluster analysis evidenced the presence of a structured diversity. In the model-based analysis, six to eight main distinct subpopulations were identified based on the molecular data. Only a relatively small portion (20%) of the molecular variation was accounted for by the geographical origin of the accessions. Major differences were detected between the North American and the Mediterranean cvs., while a considerable overlap characterized the cvs. from CIMMYT-ICARDA and Italy. The North American cvs. showed the highest within group mean genetic similarity ($GS_m = 0.68$). French cvs. revealed sizeable similarities with both the North American as well as the Italian and CIMMYT-ICARDA pools. Considering the germplasm as a whole, high levels of LD were found both at locus pairs with an intrachromosomal distance < 50 cM as well as at those with distances more than 50 cM and independent (86, 52 and 54% of SSR pairs at $p < 0.01$, respectively). After re-evaluating LD within each of the three main subgroups identified through the analysis of the germplasm structure, the LD level remained high for tightly to moderately linked locus pairs (< 20 cM apart), but was greatly reduced in the loosely linked (more than 50 cM apart) and independent locus pairs. The implications of these findings as to the possibility of using association mapping for gene/QTL discovery in durum wheat are discussed.

Introduction

Germplasm collections and breeders' materials are valuable sources for mapping genes/QTLs (Quantitative Trait Loci) through linkage analysis and/or association mapping (Jannink et al. 2001; Buckler and Thornsberry 2002; Flint-Garcia et al. 2003; Jansen et al. 2003; Rafalski and Morgante 2004). In sexually propagated species, collections of homozygous accessions are particularly suitable for the exploitation of association mapping because they allow for multiple tests over years and environments; moreover, a collection can be investigated for a wide range of traits (Morgante and Salamini 2003).

Assessing the relatedness among accessions is an important prerequisite for the identification of core germplasm collections suitable for optimizing

association studies (Garris et al. 2003; Liu et al. 2003). Depending on the average level of linkage disequilibrium (LD) present in the targeted collections and the complexity of relationships among accessions, the deployed approaches are based on the analysis of sequence haplotypes at selected candidate genes or on the whole genome scan (Thornsberry et al. 2001; Rafalski and Morgante 2004). Co-ancestry and population structure analyses, conducted on a genome-wide level using highly informative, well-distributed markers provide valuable information for association studies (Pritchard and Rosenberg 1999; Falush et al. 2003; Flint-Garcia et al. 2003). In fact, the characterization of population structure within germplasm collections is critical to identify and correctly interpret the associations between functional and molecular diversity (Pritchard and Rosenberg 1999; Buckler and Thornsberry 2002).

The presence of population structure has been widely documented in most of the studies investigating the diversity of elite crop germplasm, especially in self-pollinating cereals (Melchinger et al. 1994; Barrett and Kidwell 1998; Huang et al. 2002). The presence of distinct subgroups within a germplasm collection of a particular crop is a consequence of its prevailing mating habit (selfing vs. outcrossing), the geographic origin of the accessions, human- and environmentally driven selection, migration, genetic drift and/or the so-called founder effect (Pritchard and Przeworski 2001; Buckler and Thornsberry 2002; Flint-Garcia et al. 2003).

Durum wheat (*Triticum durum* Desf.), a selfing species, has undergone strong selection pressures throughout its breeding history (Autrique et al. 1996; Pecetti and Annicchiarico 1998). It is thus expected that, on average, the elite germplasm has a high level of LD. The medium to high level of co-ancestry among durum wheat accessions could thus be beneficially exploited in genome-wide LD association studies (Nordborg et al. 2002; Rafalski and Morgante 2004), provided that the presence of germplasm structure which could influence the average LD among unlinked markers is appropriately accounted for.

The suitability of SSR markers for evaluating the genetic relationships in the germplasm of the main crops (Huang et al. 2002; Matsuoka et al. 2002b; Garris et al. 2003; Liu et al. 2003; Thiel et al. 2003), including durum wheat (Eujayl et al. 2002), is well documented. In a previous study, SSRs were used to investigate genetic diversity among 58 accessions of durum wheat, including modern Italian cultivars and a number of important founders (Maccaferri et al. 2003). The same study also identified a set of 70 highly informative SSRs covering the whole genome and thus well-suited for further studies on the relatedness of durum wheat accessions. This set of SSR markers was utilized in the present study in order to: (i) investigate the genetic relationships of the main durum wheat gene pools cultivated world-wide, and (ii) assess, on a genome-wide scale, their level of LD and LD decay rate.

Materials and methods

Plant materials

For this study, 134 durum wheat accessions (mainly elite cultivars), representing a large portion of the genetic diversity present in the most important improved durum wheat gene pools, were considered. In particular, the accessions can be grouped into six main gene pools according to their origin as follows: group (i): 39 accessions selected and released in Italy (Italian group); group (ii): 23 hallmark accessions derived from the CIMMYT-ICARDA breeding program and released in Mexico, Spain, Italy and in several WANA (West Asia and North Africa) countries (CIMMYT-ICARDA group); group (iii): 19 accessions released by French breeders and well adapted to a range of environments throughout the centre and the south of Europe (French group); group (iv): 11 accessions derived from the Austrian or Australian breeding programs (Austrian–Australian group); group (v): 27 accessions selected in the Great Plains of the US and Canada (North American group); group (vi): 11 accessions representative of the germplasm cultivated in the southwestern region of the US under irrigation and commonly referred to as 'desert durums' (southwestern US group). In addition, four pure lines (Haurani, Inrat 69, Russello SG7 and Saragolla) selected from landraces were considered to represent the native Mediterranean germplasm.

Genotypes included in groups (i) and (ii) represent a wide range of germplasm well-adapted to the Mediterranean basin; several important

founders of the modern durum wheat gene pools grown in this area were also included. Genotypes selected in Austria or Australia have been located in the same group (iv) on the basis of their pedigree records. Most of the materials included in groups (i), (ii) and (vi) are semi-dwarf, day-length insensitive (medium to early flowering) genotypes, while some of the French genotypes, group (iii), and the majority of the North American, group (v), are day-length sensitive genotypes. The main details on the 134 accessions, 58 of which have also been considered in the study of Maccaferri et al. (2003), have been reported in Table 1.

Based on the most recent durum wheat diffusion and production data, among the 134 accessions herein considered we have identified a subset of 93 cvs. (Table 1) which account for most of the genetic diversity presently available within the cultivated materials of the six main gene pools herein investigated.

Molecular data

For each accession, DNA was extracted from the leaves of 20 plantlets (Saghai-Maroof et al. 1984). The same set of SSR markers investigated by Maccaferri et al. (2003) was utilized in the present study. The 70 SSRs [69 dinucleotide genomic loci and one trinucleotide locus developed by Röder et al. (1998) and Devos et al. (1995), respectively] were selected for their reliability, level of polymorphism and genome coverage. Amplification products were obtained with minor modifications to the protocol of Röder et al. (1998). Highly polymorphic dinucleotides were preferred over tri- and tetra-nucleotides in order to limit the risk of allelic homoplasy (presence of alleles similar in state, but not identical by descent; Estoup et al. 2002), a condition less likely to occur with dinucleotide SSRs (Matsuoka et al. 2002a,b).

Data analysis

Pairwise genetic similarity values (GS_{ij}) among all accessions were calculated as the proportion of loci with shared alleles (Lu and Bernardo 2001; Matsuoka et al. 2002a) in NTSYS-pc software version 2.0 (Rohlf 1997), by using the simple matching coefficient for multi-state qualitative data:

$$GS_{ij} = m/n$$

where m is the number of loci with allelic variants of the same molecular weight present in the two accessions i and j being compared, summed over all the surveyed loci, and n is the total number of loci, excluding loci with missing data.

The diversity index of each microsatellite locus was calculated as:

$$GD = 1 - \Sigma p_j^2$$

where p_j is the frequency of the jth allele across all accessions (Powell et al. 1996).

Cluster analysis using a distance-based method (SAHN method, UPGMA algorithm) of the 134 accessions and principal coordinate analysis (PCoA) of the subset of 93 representative modern cvs. were carried out by using the genetic similarity matrix (Gower 1972). Analyses were carried out with the NTSYS-pc software.

Considering the subset of 93 cvs., AMOVA (Analysis of Molecular Variance; Excoffier et al. 1992) was used to test the significance of the partitioning of genetic variance among the six main gene pools. Pairwise F_{st} genetic distances were also computed as measures of the genetic diversity between gene pools (Reynolds et al. 1983; Slatkin 1994). Calculations were carried out in ARLEQUIN version 2.0 (Schneider et al. 2000).

The genetic diversity structure of the entire germplasm herein analysed (134 accessions) was also investigated with an alternative approach by using the model-based (Bayesian) clustering algorithm (STRUCTURE software; Pritchard et al. 2000) which identifies subgroups of accessions with distinct allele frequencies within the germplasm. Differently from the cluster analysis, which is based on the calculation of a pairwise distance matrix and on a 'non-overlapping' graphical representation, in the model-based method each accession is allowed to have membership in several different subgroups, with membership coefficients totalling 1. The program was run for a number (K value) of hypothetical subgroups ranging from two to eight. Runs were carried out by setting for 100,000 iterations, of which only the last 50,000 were recorded, and assuming an admixture linkage model with correlated allele frequencies (Falush

Table 1. List of the 134 durum wheat accessions (cultivars, landrace selections and breeding lines) and registration details

Genotype	Registration		Pedigree	Seed source[a]	Sel. cvs[b]
	Country	Year			
Adamello	Italy	1985	Valforte/Turkish selection	1	
Appio	Italy	1982	Cappelli//Gaviota/Yuma	2	x
Appulo	Italy	1973	Cappelli/Grifoni//Capeiti 8	1	x
Arcangelo	Italy	1983	Creso/Appulo	1	x
Bronte	Italy	1996	Berillo/Latino	2	x
Capeiti 8	Italy	1940	Cappelli/Eiti	3	
Cappelli	Italy	1930	Strampelli' selection from Jennah Khetifa	1	
Ciccio	Italy	1996	F6 Appulo/Valnova//Valforte/Patrizio	4	x
Cirillo	Italy	1992	Jucci/Polesine//Creso/Montanari	4	
Colosseo	Italy	1995	Mexa's mutant/Creso	4	x
Creso	Italy	1974	Yt 54-N10-B/2*//3*TC 60/3/Cp B 14	4	x
Duilio	Italy	1984	Cappelli//Anhinga/Flamingo	2	x
Flaminio	Italy	1998	Latino/Cappelli	2	
Flavio	Italy	1992	Latino/Cappelli	2	
Fortore	Italy	1995	Capeiti 8/Valforte	1	
Gargano	Italy	1997	Trinakria/Valforte//Valnova/Appulo	1	
Grazia	Italy	1985	M 6800127/Valselva	4	x
Ionio	Italy	1995	Lira/Vic	4	x
Iride	Italy	1996	Altar 84/Ares = Ionio	4	x
Italo	Italy	1993	Cross between Italian and Turkish cvs.	4	x
Karel	Italy	1980	Mex.198/Maristella	5	
L35	Italy	n.a.	Breeding line derived from Altar 84/Ares	4	
Latino	Italy	1982	Cappelli/Aningha//*T. turgidum*	2	
Lira B 45	Italy	1985	Mandon/FD 1104	4	
Messapia	Italy	1982	Mex./Crane "S"//Tito	4	
Ofanto	Italy	1990	Appulo/Adamello	4	x
Platani	Italy	1995	Valnova/Capeiti	5	
Plinio	Italy	1988	Linea D50/Trigo Candeal	2	
San Carlo	Italy	1996	Grazia/Degamit	4	
Produra	Italy	1980	TME/2*TC60//Wells/3/TC60/2*BYE//Tecur125E/*TC60	5	
Simeto	Italy	1988	Capeiti 8/Valnova	4	x
Solex	Italy	1995	Creso/Valgerardo	4	
Svevo	Italy	1996	CIMMYT's Selection/Zenit	4	x
Trinakria	Italy	1970	B 14/Capeiti 8	4	
Valbelice	Italy	1992	0111/BC 5	1	x
Valforte	Italy	1980	Yt54-N10B/2*BY//LD390 II 14587/3/Cappelli*2/Yuma	5	
Valnova	Italy	1975	Yt54-N10B/2*BY//LD390 II 14587/3/Cp/4/Cp/Yuma	1	
Varano	Italy	1997	Capeiti 8/Creso//Creso/3/Valf./Trinakria	4	
Zenit	Italy	1992	Valriccardo/Vic	4	
Russello SG7	Italy	n.a.	Landrace selection, from "Russie"	3	
Saragolla	Italy	n.a.	Landrace selection, from "Saragolle"	3	
Haurani	Syria	n.a.	Local landrace selection	11	
Inrat 69	Tunisia	1969	Mahamoudi/Kyperounda	14	
Acalou	France	1990	Valsacco/Ranger	6	x
Arcalis	France	1995	Edmore/Creso	6	x
Agridur	France	1988	Edmore//CIMMYT 303/Chandur	6	x
Ardente	France	1984	Israel durum 303/preliminary77//664	6	x
Aramon	France	1987	1971-4/r	6	x
Arstar	France	1992	Valselva/71-75	6	x
Brindur	France	1987	Crosby/623//Edmore	6	x
Durfort	France	1996	Selected from Reva population	6	x
Duriac	France	1991	n.a.	6	x
Excalibur	France	1990	646/Mondur	6	x
Exeldur	France	1992	Valdur/Regal	4	x
Galadur	France	1992	Edmore/Blondur//Montferrier/DT192	6	x

Table 1. Continued

Genotype	Registration		Pedigree	Seed source[a]	Sel. cvs[b]
	Country	Year			
Ixos	France	1990	Valnova/3/Tomclear/662//662	4	x
Orjaune	France	1995	miradur/idyn81-04	6	x
Nefer	France	1996	164/Keops	4	x
Neodur	France	1987	184-7/Valdur//Edmore	4	x
Primadur	France	1984	Blondur//2587/Leeds	6	x
Tetradur	France	1992	Edmore//Capdur/Regal	6	x
Auroch	France	1997	Blé dur N.Dakota n ° 79168	6	x
AC Avonlea	Canada	1997	DT379/DT367//DT367/Medora	7	x
AC Melita	Canada	1994	Medora/Lloyd	7	x
AC Morse	Canada	1996	RL 7196/D84328	7	x
AC Navigator	Canada	1998	Kyle/WB881	7	x
AC Pathfinder	Canada	1998	DT367/WB881	7	x
Hercules	Canada	1969	Ld308/Ld368//Stewart/Ld393	7	
Kyle	Canada	1984	Wakooma/DT320//Wakooma/DT322	7	x
Medora	Canada	1982	Ward/Macoun	7	x
Plenty	Canada	1990	Vic/Waskana//Hercules/DT310	7	x
Sceptre	Canada	1985	D72110/Coulter	7	x
Wakooma	Canada	1973	Lakota*2/Pelissier	7	
Waskana	Canada	1970	Lakota*2/Pelissier	7	
Ben	USA[c]	1996	D8024/Monroe	8	x
Belzer	USA[c]	1997	D7798/DT367	8	x
Plaza	USA[c]	1999	Plenty/D8291	8	x
Lloyd	USA[c]	1983	Cando/Edmore	8	x
Maier	USA[c]	1998	D8193/D8335	8	x
Monroe	USA[c]	1985	D7456/Vic	8	x
Munich	USA[c]	1995	D8030/D8016	8	x
Renville	USA[c]	1988	Rolette/Vic	8	x
Rugby	USA[c]	1974	Langdon/3/Ld357//CI7780/Ld362/4/Br180/Wells	8	x
Lakota	USA[c]	1960	Sentry//Ld379/Ld357	8	
Langdon	USA[c]	1956	Mindum/Carleton//Khapli/3/Heiti/Stewart//Mindum/ Carleton/4/Stewart/5/Carleton	8	x
Edmore	USA[c]	1978	D6530/D65114	8	x
Vic	USA[c]	1979	Edmore/Ward	8	x
Yuma	USA[c]	1956	Ld194/Khapli emmer/3/Ld308	8	
Mindum	USA[c]	1917	farmer selection	8	
West Bred 881	USA[d]	n.a.	Cross among Ward Wells Cando Waskana Mexicali 75 1000D	9	x
Kronos	USA[d]	n.a.	APB MSFRS POP Sel D03-21	4	x
Colorado	USA[d]/Italy	1995	P 92/932-2	4	x
West Bred Turbo	USA[d]	1985	Cross among Italian and CIMMYT cvs.	9	x
Tacna	USA[d]	n.a.	Durum S-1/E. Wheat −89 S-1	9	x
Mohawk	USA[d]	n.a.	883-22 Alpha Pop - 85 CHA	9	x
Cortez	USA[d]	n.a.	Turbo Alpha Pop - 86 CHA	9	x
Bravadur	USA[d]	n.a.	WWW MSFRS Pop	10	x
Reva	USA[d]	n.a.	WWW MSFRS Pop	10	x
Durex	USA[d]	n.a.	WWW MSFRS Pop	10	x
Duraking	USA[d]	n.a.	WWW MSFRS Pop	10	x
Anton	Spain	1991	n.a.	13	x
Arcobaleno	Spain/Italy	1995	Chen/Altar 84	4	x
Don Pedro	Spain	n.a.	CARC/AUK	13	x
Jabato	Spain	1989	n.a.	11	x
Roqueno	Spain	1991	n.a.	11	x
Vitromax	Spain/Italy	1996	Turchia77/3/Jori/Anhinga//Flamingo	1	
Vitron	Spain/Italy	1987	Turchia77/3/Jori/Anhinga//Flamingo	1	x
Belik 2	Lebanon	1987	CR/STK	11	x

Table 1. Continued

Genotype	Registration		Pedigree	Seed source[a]	Sel. cvs[b]
	Country	Year			
Kabir 1	Algeria	1993	OVI/CP/2FG	11	x
Karim	Tunisia	1983	Jori"S"/Anhinga"S"//Flamingo"S"	14	
Khiar	Tunisia	1983	Chen/Altar84	14	
Korifla Cham 3	Syria	1987	DS15/GEIER	11	x
Omrabi 3 Cham 5	Syria	1993	JO/Haurani	11	x
Heider	Syria	1997	CAN2109/2/JO/AA/3/S15/CR	11	x
Waha Cham 1	Syria	1984	PLC/RUF/2/GTA/RTTE	11	x
Altar 84	Mexico	1984	Ruff "S"/Flamingo "S"//Mexicali 75/3/SHWA"S"	4	x
Aconchi 89	Mexico	1989	Altar 84/Araos	4	x
Cocorit71	Mexico	1970	RAE/4*TC60//STW63/3/AA"S"	12	
Crane	Mexico	n.a.	BYE*2/TC60//STW63/3/ZB/WLS/4/GLL"S"	12	
Gaviota	Mexico	1972	CR"S"/4/T.POL.185309//T.GLE/2*TC60/3/GLL"S"	12	
Jori C 69	Mexico	1969	BYE*2/TC60//TAC125E/3*TC60	12	
Gallareta	Mexico	1982	RUFF"S"/FG"S"//MEXI75/3/SHWA"S"	12	
Mexicali 75	Mexico	1975	61.130/Leeds//Jori"S"/3/GDOVZ469	4	x
Astrodur	Austria	1991	Valdur//Pandur/Valgerardo	15	x
Extradur	Austria	1993	Mondur/Grandur//Astrodur	15	x
Goldur	Austria	1989	Valdur//Pandur/Valgerardo	15	x
Grandur	Austria	1980	Adur/unknown	15	x
Frankodur	Austria	1999	Mondur/Grandur//Astrodur	15	x
Helidur	Austria	1994	Signadur/Astrodur	15	x
Semperdur	Austria	1996	Astrodur/Kamilaroi	15	x
Topdur	Austria	1995	Kamilaroi/Astrodur	15	x
Kamilaroi	Australia	1982	Durati/Leeds	4	x
Wallaroi	Australia	n.a.	TAMB-17/Kamilaroi	4	x
Yallaroi	Australia	n.a.	n.a.	4	x

[a]Seed Sources: 1 – Ente Nazionale Sementi Elette (ENSE), Milano, Italy; 2 – Società Italiana Sementi (SIS), Bologna, Italy; 3 – Istituto del Germoplasma, Bari, Italy; 4 – Società Produttori Sementi Bologna (SPB), Bologna; 5 – Ist. Sper. Cerealicoltura, Sezione di Foggia, Foggia, Italy; 6 – Groupe d' Etude et de controle des Variétés et des Semences (GEVES), GEVES La Minière, Guyancourt Cedex, France; 7 – Agriculture and Agri-Food Canada Semiarid Prairie Agriculture Research Centre (AAFC SPARC), Swift Current, SK, Canada; 8 – North Dakota State University (NDSU), Fargo, North Dakota, USA; 9 – Western Plant Breeder (WPB), Bozeman, Montana, USA; 10 – World Wide Wheat (WWW), Phoenix, Arizona, USA; 11 – ICARDA: International Centre for Agricultural Research in the Dry Areas, Aleppo, Syria; 12 – JIC: John Innes Centre, Norwich, UK; 13 – UdL-IRTA: Institute of Agro-food Research and Technology IRTA and University of Lleida, Lleida, Spain; 14 – INRAT, Institut Nazionale de la Recherce Agricole, Ariana, Tunisia; 15 – Probstdorfer Saatzucht, Probstdorfer, Austria.

[b]Modern cultivars, 93 in total, chosen as representative of the genetic diversity present in the six main elite durum wheat gene pools herein considered.

[c]North Dakota.

[d]Southwestern USA.

et al. 2003). No a priori population information was used. Correlation values were computed using the results of five independent runs of STRUCTURE carried out for each of the tested K values.

LD estimate and significance for each pair of SSR loci were evaluated by using the software package TASSEL (www.maizegenetics.net/bioinformatics/tasselindex.htm). LD was evaluated for the entire population considered and for selected subgroups as identified by the model-based subdivision. SSRs were filtered for containing no more than 5% of missing data (including null alleles). Due to the high frequency of rare alleles at many SSRs, for each SSR the alleles with very low (< 0.07) overall frequencies were pooled in a common allelic class, thus limiting the inflation effect of rare alleles on LD estimates and in particular on p-value (Mohlke et al. 2001; McRae et al. 2002). Six SSRs (*Xgwm6*, *Xgwm136*, *Xgwm247*, *Xgwm302*, *Xgwm448* and *Xgwm544*) with a total frequency of pooled rare alleles higher than 20% were discarded. In total, 58 out of the 70 tested SSRs met the required conditions. D' and r^2 LD measures modified for loci

with multiple alleles were used (Hedrick 1987; Farnir et al. 2000). Significance (p values) of LD for SSR pairs was determined by permutation (Weir 1996); for each pair, 100,000 permutations were performed.

The map position for most of the considered SSR loci was inferred from Röder et al. (1998); a few additional loci were positioned according to the results of Peng et al. (2000) and Nachit et al. (2001).

Results

The total number of alleles amplified at the 70 SSR loci in the 134 accessions was equal to 485 (6.9 alleles per locus on average, ranging from 2 to 17); the diversity index ranged from 0.09 to 0.81 (mean value of 0.55).

Genetic diversity within and among the elite durum wheat gene pools

In order to assess the genetic diversity present within and among the main groups, 93 cvs. representing the six main gene pools (see Table 1) were chosen based on their origin and relative diffusion. In this subset of leading cvs., 389 alleles (5.5 alleles per locus) were detected, i.e. a 19.8% reduction as compared to the total number of alleles detected in the complete population of accessions. However, the mean SSR diversity content in this subsample was similar to that of the whole population (mean diversity index equal to 0.53 and 0.55, respectively).

PCoA was used to depict the relationships among the 93 leading durum cvs. (Figure 1). The first two principal components accounted for 20% of the total variance. The most diverse cvs. (outermost entries) of each of the six origin-based groups have been connected by lines that identify convex hulls representing the genetic diversity present within each group. A major diversification was observed between the Mediterranean gene pool (comprising the Italian pool, the CIMMYT-ICARDA pool and some French cvs.; quadrants III and IV) and the North American pool (comprising the modern cvs. from Canada and North Dakota; quadrants I and II). The Italian group accounted for a sizeable portion of the total genetic diversity (cvs. spread in most of quadrants III and IV) and extensively overlapped with the CIMMYT-ICARDA derived cvs. The modern cvs. from North America appeared to be the least diversified, while the group of recent French cvs. spanned most of the genetic diversity space represented in the PCo diagram and showed relationships with both the North American and the Mediterranean pools. This finding indicates that French breeders have been very effective in exploiting the genetic diversity of different gene pools, despite the short history of durum wheat breeding in France. Early flowering, day-length insensitive durums from Arizona and California (southwestern US group), positioned in the centre of the PCo diagram, were also well-diversified. Although the Austrian–Australian group accounted for a limited portion of the genetic diversity, its centroid is distinct from those of the other groups and the third principal coordinate, accounting for 5% of the total variance, clearly discriminated this group from the others (data not shown).

The structure of the genetic diversity based on the different origins of the 93 cvs. was tested by AMOVA. Although the difference among groups was significant ($p \leq 0.001$), the within group component of variance prevailed (79.5% of the total variation). Table 2 reports the pairwise F_{st} genetic distances between groups and a summary of allele diversity statistics for each group. These data confirmed the relationships among groups represented by PCo analysis. The lowest degree of diversity was observed between the pools from Italy and CIMMYT-ICARDA ($F_{st} = 0.07$). As compared to the other groups, the North American pool showed the highest genetic distance values (F_{st} ranging from 0.30 to 0.33), except with the French materials ($F_{st} = 0.16$); on the contrary, French cvs. showed the lowest pair-wise F_{st} values with all groups (ranging from 0.11 to 0.19). As to the within-group mean genetic similarity, the North American group had the highest GS_m value (0.68). Among groups, the mean number of alleles per locus (a raw index of diversity) was relatively high for the Italian, French and CIMMYT-ICARDA groups (3.7, 3.5 and 3.1 alleles/locus, respectively), while the number of alleles specific of each group was high only for the Italian and French groups (42 and 26 alleles, respectively).

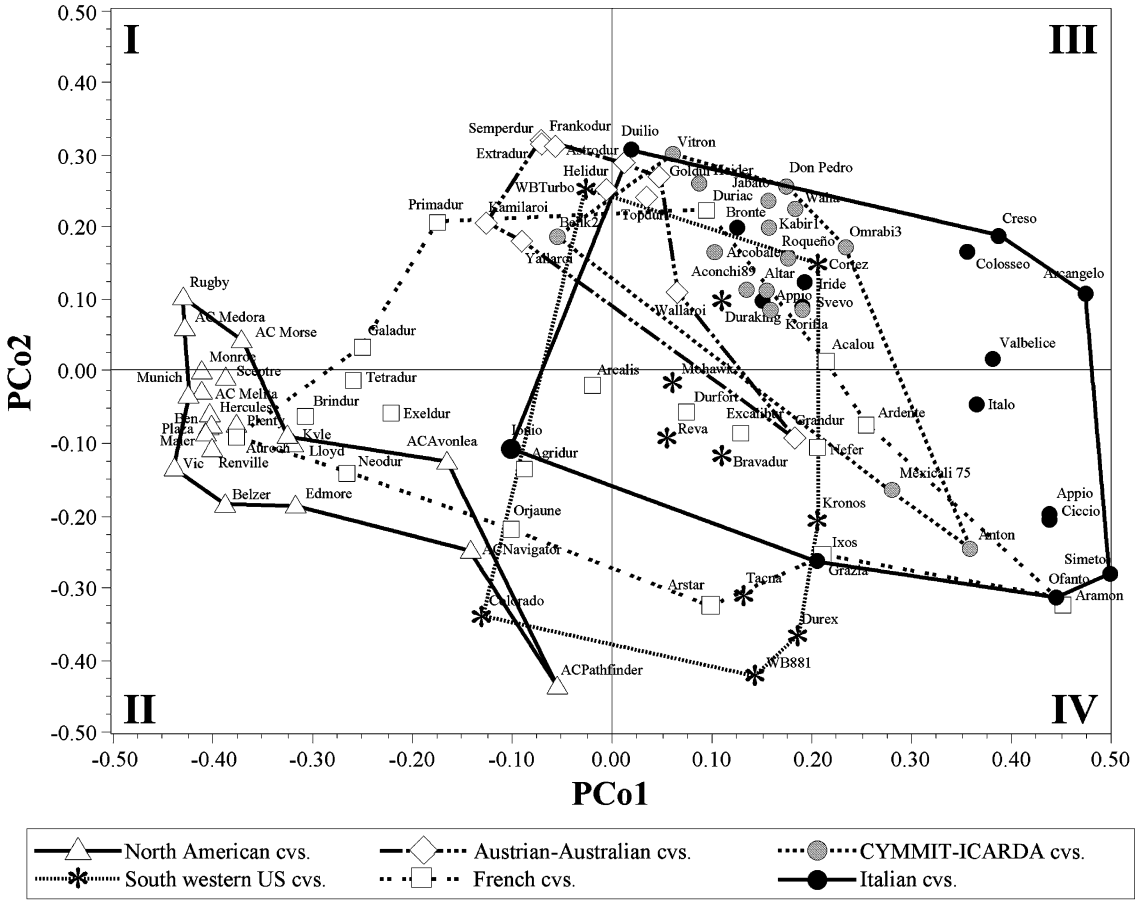


Figure 1. Principal coordinate analysis of 93 elite cultivars (cvs.) based on the analysis of 70 SSRs. Genetic similarities among cvs. have been calculated as the proportion of loci with shared alleles. The cvs. have been chosen to represent the genetic diversity present in six main gene pools identified by origin (North America, southwestern US, Austria and Australia, France, Italy and the CIMMYT-ICARDA breeding program). To represent the portion of genetic variability accounted for by each gene pool, the most diverse cvs. of each group have been connected by lines.

Table 2. Genetic diversity statistics of the six groups: mean genetic similarity (GS_{ij}) within-group, pair-wise F_{st} genetic distance between groups, mean number of different alleles per SSR locus and total number of group-specific alleles, summed over the 70 surveyed SSRs

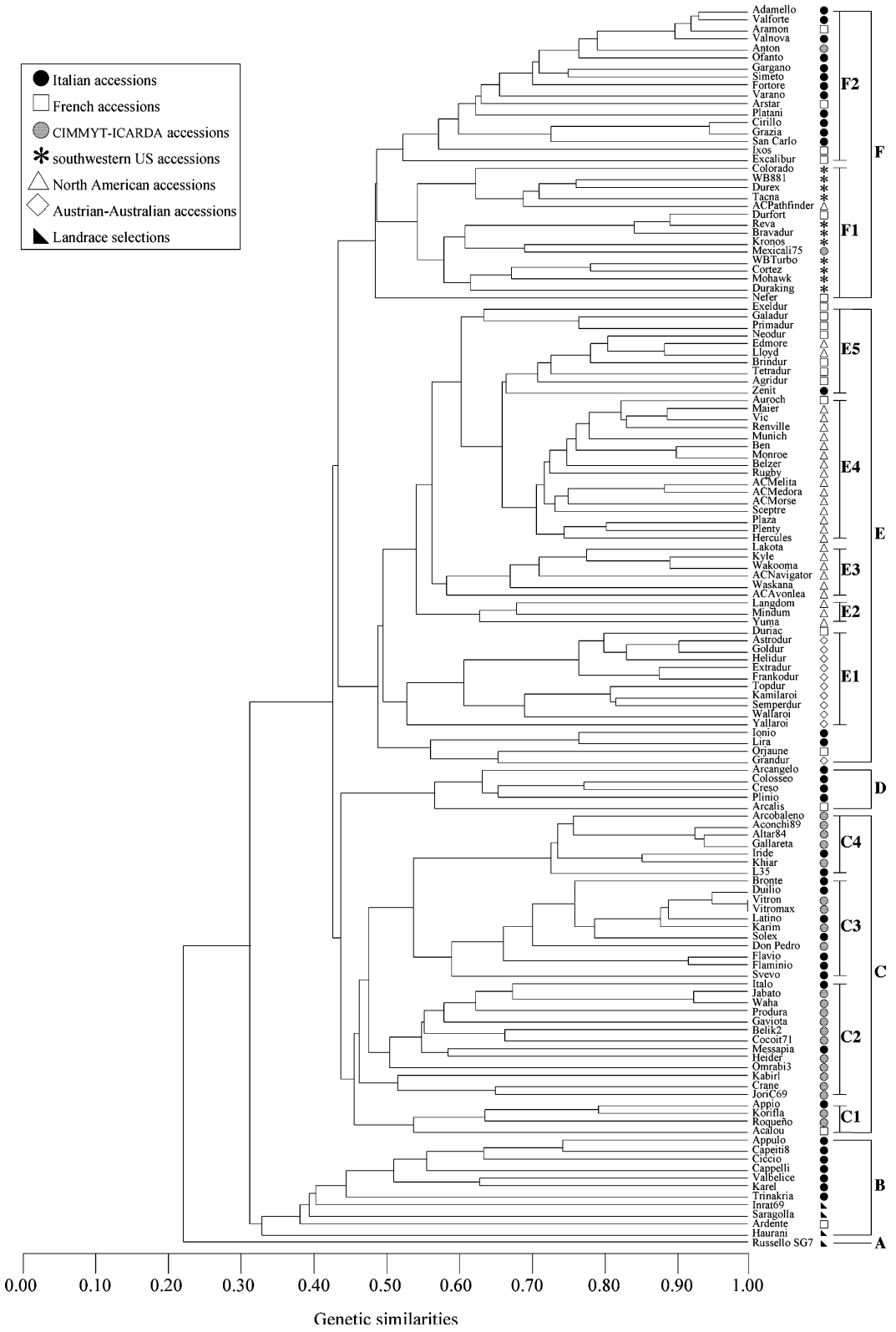
Origin	Sample size	Italy	CIMMYT-ICARDA	South western US	France	North America	Austria–Australia
Italy	16	**0.43**[a]					
CIMMYT-ICARDA	15	0.07**[b]	**0.51**				
South-western US	11	0.17***	0.15***	**0.59**			
France	19	0.11***	0.12***	0.12***	**0.50**		
North America	21	0.31***	0.30***	0.30***	0.16***	**0.68**	
Austria–Australia	11	0.23***	0.29***	0.29***	0.19***	0.33***	**0.64**
Alleles/locus (no.)		3.7	3.1	2.5	3.5	2.5	2.9
Group-specific alleles (no.)		42	11	7	26	18	12

[a]Mean genetic similarity (GS_{ij}) within-group (in bold).
[b]Pairwise F_{st} genetic distance between groups.
, *: Significant at $p < 0.01$ and $p < 0.001$, respectively.

Distance-based and model-based analysis of the genetic relationships in the complete set of accessions

The distance-based cluster analysis (Figure 2) shows a main subdivision of accessions in at least five quite distinct groups. The accession Russello SG7 (*Triticum durum* sect. *elongata* Vav. var. *hordeiforme* Hust), a selection from a population native to Sicily (southern Italy) shows no appreciable genetic similarity to any of the modern elite durum wheat breeding lineages of different origin. In Figure 2, Russello SG7 has been highlighted with the capital letter 'A' and the five major groups, which include most of the analysed accessions, have been identified by capital letters from 'B' to 'F'. A primary distinction separates a group of accessions all belonging to the old germplasm native to the Mediterranean basin (group B) from most of the modern improved germplasm; group B includes old accessions selected from landraces or from crosses between landrace-derived lines. Genealogies of these accessions include both representatives of the North African germplasm (e.g., Cappelli, Inrat 69 and Saragolla) and of the west Asian germplasm (e.g., Haurani and genotypes related to Eiti). The modern germplasm is divided in four main groups and several subgroups (indicated in Figure 2 by combinations of letters and numbers) that identify distinct breeding lineages; this classification is largely in accordance with the available pedigree information. In the lower part of the dendrogram, group C includes various materials comprising hallmark CIMMYT-ICARDA genotypes and successful CIMMYT-derived cultivars selected within different national breeding programs. Within group C, quite distinct breeding lineages are those referring to the CIMMYT founders 'JO/AA//FG-cross' (subgroup C3) and 'Altar 84' (subgroup C4), while several old, 'first generation' CIMMYT-ICARDA materials are clustered in subgroup C2. The group D includes Italian cvs. obtained from the founder Creso. The main group E coincides with materials tracing back mainly to the gene pool originally developed in North America. Within this group, a number of clusters of related accessions, mostly in accordance with the expected co-ancestry data, were identified as follows: subgroup E1 (the most distinct) with the Austrian–Australian accessions; subgroup E2 with the old North American founders; subgroup E3 with the Canadian accessions derived from the US founder 'Lakota'; subgroup E4 with recent cvs. from North Dakota and Canada; subgroup E5 with French cvs. derived from the North Dakota founders. Similarly to PCoA, cluster analysis pointed out that the subgroup E4 of elite modern US and Canadian cvs. is characterized by a high level of genetic similarity. In the upper part of the dendrogram, subgroup F2 includes most of the modern Italian cvs. derived from the early CIMMYT-derived founders Valnova and Valforte, while subgroup F1 includes the 'desert durums', partially related to the successful CIMMYT founder Mexicali 75, in turn derived from a Valnova sib. A more detailed analysis of the genetic relationships among the materials released in Italy has been reported in Maccaferri et al. (2003).

The model-based Bayesian cluster analysis was performed using all the 134 accessions and five independent runs of STRUCTURE for each K value (hypothetical number of subpopulations) from 2 to 8. For each K value, the run showing the highest posterior probability of data was considered. The Bayesian posterior probability of data steadily improved until $K = 4$ and, to a lower extent, until $K = 8$. However, the highest number of accessions assigned to a specific cluster with a probability higher than 80% was obtained with $K = 6$ (75 accessions, i.e., 56% of the total), while with $K = 7$ and 8 this percentage dropped to 34% of the total number of accessions, thus indicating the presence of complex relationships among accessions. The clustering diagrams with K ranging from 2 to 6 are reported in Figure 3. The main subdivision between the photoperiod-insensitive (adapted to the temperate, Mediterranean conditions) and the photoperiod-sensitive (mainly derived from the North American materials) gene pools is evident with $K = 2$. The three subpopulations identified by structure analysis ($K = 3$) largely correspond to gene pools originated from North America (green colour in Figure 3; subpop. A in Figures 4 and 5), CIMMYT-ICARDA (red colour in Figure 3; subpop. B in Figures 4 and 5), and Mediterranean basin (yellow colour in Figure 3; subpop. C in Figures 4 and 5). When considering four subpopulations, the less complex structure of the North American-related genotypes, as compared to that of the main group of


Italian accessions
French accessions
CIMMYT-ICARDA accessions
southwestern US accessions
North American accessions
Austrian-Australian accessions
Landrace selections
Adamello
Valforte
Aramon
Valnova
Anton
Ofanto
Gargano
Simeto
Fortore
Varano
Arstar
Platani
Cirillo
Grazia
San Carlo
Ixos
Excalibur
Colorado
WB881
Durex
Tacna
ACPathfinder
Durfort
Reva
Bravadur
Kronos
Mexicali75
WBTurbo
Cortez
Mohawk
Duraking
Nefer
Exeldur
Galadur
Primadur
Neodur
Edmore
Lloyd
Brindur
Tetradur
Agridur
Zenit
Auroch
Maier
Vic
Renville
Munich
Ben
Monroe
Belzer
Rugby
ACMelita
ACMedora
ACMorse
Sceptre
Plaza
Plenty
Hercules
Lakota
Kyle
Wakooma
ACNavigator
Waskana
ACAvonlea
Langdom
Mindum
Yuma
Duriac
Astrodur
Goldur
Helidur
Extradur
Frankodur
Topdur
Kamilaroi
Semperdur
Wallaroi
Yallaroi
Ionio
Lira
Orjaune
Grandur
Arcangelo
Colosseo
Creso
Plinio
Arcalis
Arcobaleno
Aconchi89
Altar84
Gallareta
Inde
Khiar
L35
Bronte
Duilio
Vitron
Vitromax
Latino
Karim
Solex
Don Pedro
Flavio
Flaminio
Svevo
Italo
Jabato
Waha
Produra
Gaviota
Belik2
Cocuit71
Messapia
Heider
Omrabi3
Kabiri
Crane
JoriC69
Appio
Korifla
Roqueño
Acalou
Appulo
Capeiti8
Ciccio
Cappelli
Valbelice
Karel
Trinakria
Inrat69
Saragolla
Ardente
Haurani
Russello SG7
F2
F1
F
E5
E4
E3
E2
E1
E
D
C4
C3
C2
C1
C
B
A
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
Genetic similarities

◄

Figure 2. UPGMA dendrogram showing the pattern of genetic diversity among the 134 durum wheat accessions based on the analysis of 70 SSRs. Genetic similarities among accessions have been calculated as the proportion of loci with shared alleles. Capital letters indicate the main subgroups; combinations of letters and numbers indicate the different breeding lineages. A: the old landrace selection Russello SG7; B: old Mediterrnean germplasm; C: CIMMYT-ICARDA, Italian and Spanish accessions mainly derived from outstanding CIMMYT genotypes (C1: Korifla-group; C2: various accessions from CIMMYT-ICARDA; C3: JoriC69/AA//FG-derived group; C4: Altar 84-group); D: Creso-derived group (Italian accessions); E: North American- and Austrian–Australian-derived groups (E1: Austrian–Australian group; E2: old North American founders; E3: Canadian accessions derived from the US founder Lakota; E4: North Dakota and Canadian accessions; E5: French accessions derived from the North Dakota founders); F: Valnova-Valforte- and Mexicali 75-derived groups (F1: 'desert durums', related to Mexicali 75; F2: Italian accessions derived from Valnova-Valforte).

Table 3. Number of SSR locus pairs used to investigate the presence and level of long-range LD between SSR loci in the entire population (134 accessions) and in each of the three main subpopulations separated according to the STRUCTURE results with $K = 3$

	Polymorphic loci (no.)	Locus pairs					
		Total (no.)	With intermarker distance				Independent (no.)
			0–10 cM (no.)	11–20 cM (no.)	21–50 cM (no.)	> 50 cM (no.)	
Entire collection	58	1653	9	10	30	46	1558
North American subpopulation	55	1485	9	7	31	42	1396
CIMMYT subpopulation	55	1485	9	10	30	44	1392
Mediterranean subpopulation	55	1485	9	10	30	44	1392

Subpopulations A, B and C comprise 54 North American related accessions, 48 recent CIMMYT-ICARDA derived accessions and 32 old Mediterranean or Italian accessions, respectively. SSR locus pairs have been classified based on the intermarker genetic distance [as estimated from Röder et al. (1998)] as follows: tightly to moderately linked (0–10 and 11–20 cM apart), loosely linked (21–50 cM), with a distance > 50 cM and independent.

Mediterranean and CIMMYT-ICARDA derived accessions, becomes evident. With $K = 6$, this latter group is subdivided in four subgroups, which largely correspond to groups B, C1 and C2, C3 and C4, and F2 as identified in the distance-based dendrogram. Within the North American-derived materials, a clearly distinct subgroup of Austrian–Australian accessions clusters apart. This finding is in keeping with the results of the dendrogram reported in Figure 2. The 'desert durums' can be considered as the group of recent elite materials with the highest level of genetic complexity/admixture, among the others. With $K = 6$, most of these cvs. showed partial membership to three or four different subgroups belonging to the main groups of Mediterranean, CIMMYT-ICARDA and North American gene pools, a result in keeping with the results of PCoA (Figure 1) and the origin of these cvs., selected either from multiple crosses or through recurrent selection within populations derived from several foundation genotypes. Generally, the attribution of an accession, based on its membership coefficients, to two or more subgroups was in agreement with its known origin/pedigree data.

Level of long-range LD in the durum germplasm

The number of locus pairs obtained with the selected SSRs (see Materials and methods) and used to estimate the LD for the entire collection and for each of the three main subpopulations is reported in Table 3. The locus pairs were subdivided in five classes on the basis of their inter-marker genetic distance, i.e., tightly to moderately linked (≤10 and 11–20 cM apart), loosely linked (21–50 and > 50 cM) and independent pairs. The analysis of the entire collection was conducted using the 58 selected SSRs for a total of 1653 locus pairs, whereas, due to the monomorphism of three different SSRs within each subpopulation, the analysis of each subpopulation was based on 55 SSR markers evidencing a total of 1485 locus pairs. In all the cases the majority of locus pairs (94%) was represented by independent loci; tightly to mod-

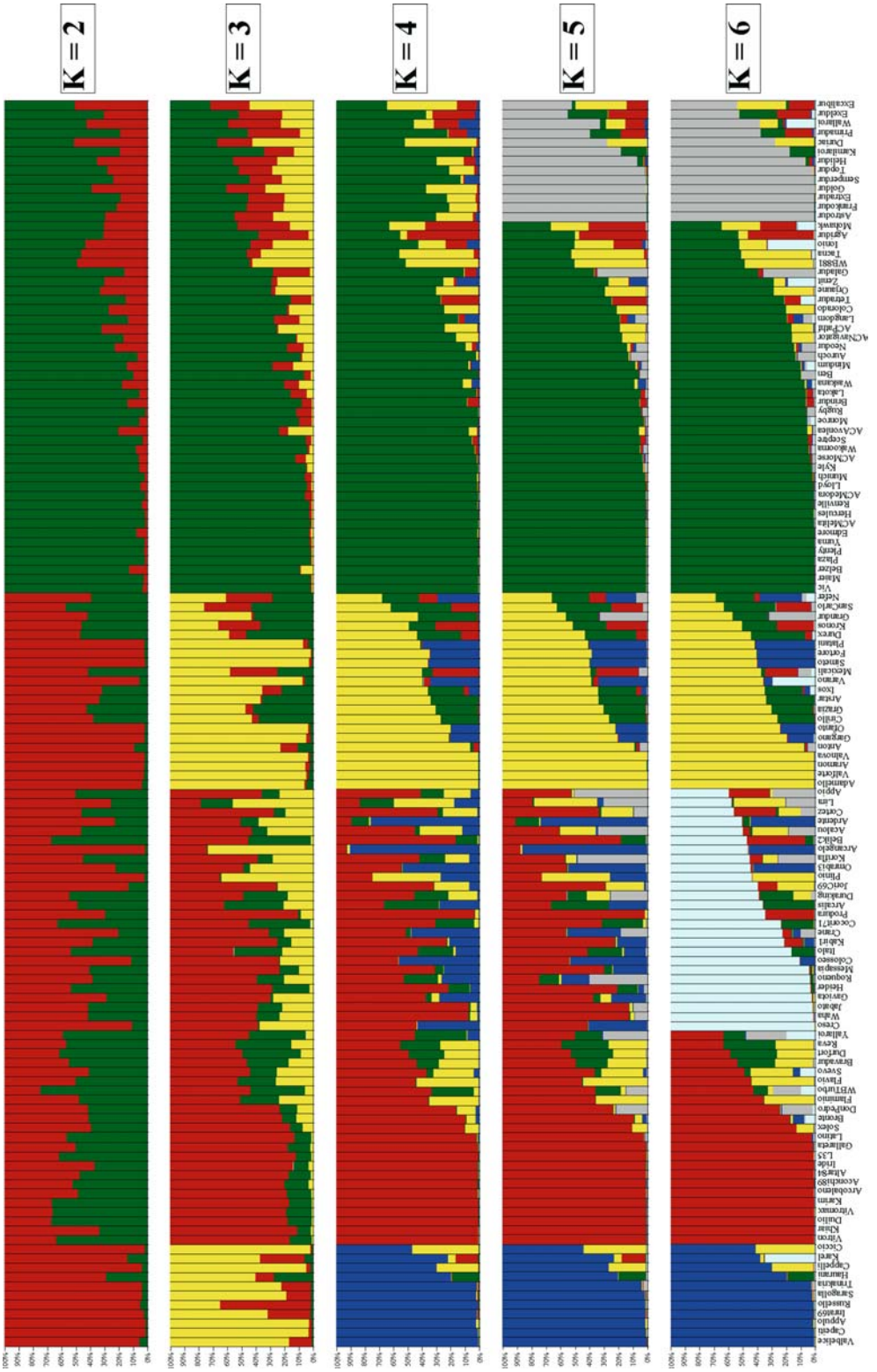


Figure 3. Genetic diversity structure of the 134 durum wheat accessions as estimated using the model-based Bayesian algorithm implemented in the program STRUCTURE. Population memberships (expressed as %) for each accession are shown as estimated using a number of hypothetical subpopulations varying from $K = 2$ to $K = 6$. The main subpopulations obtained with $K = 6$ are identified by different colours as follows: blue: old Mediterronean germplasm; red: accessions derived from recent CIMMYT outstanding genotypes; pale blue: the CIMMYT-ICARDA and Creso-derived accessions; yellow: accessions related to the Italian Valnova-Valforte founders; green: accessions mainly related to the North American gene pool; grey: Austrian–Australian accessions.

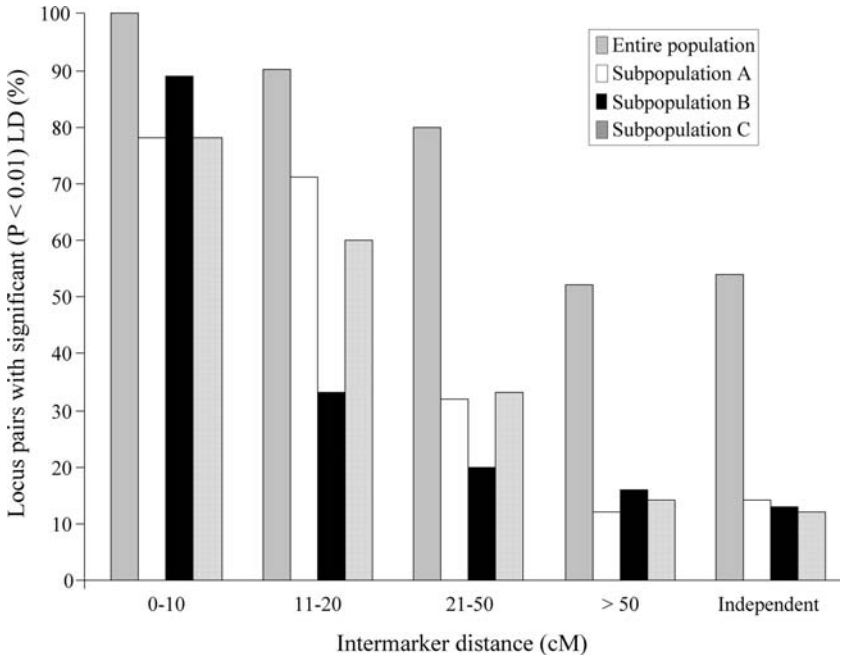


Figure 4. Average level of long-range LD between SSR loci. Pair-wise SSR markers have been classified based on the intermarker genetic distance (Röder et al. 1998) as follows: tightly to moderately linked (0–10 and 11–20 cM apart), loosely linked (21–50 cM), with a distance > 50 cM and independent. LD analysis was performed on the entire population (134 accessions) and on each of the three main subpopulations separated according to the STRUCTURE results with $K = 3$. Subpopulations A, B and C comprise 54 North American related accessions, 48 recent CIMMYT-ICARDA derived accessions and 32 old Mediterranean or Italian accessions, respectively. For each class, the percentage of locus pairs found in significant LD at $p < 0.01$ has been reported.

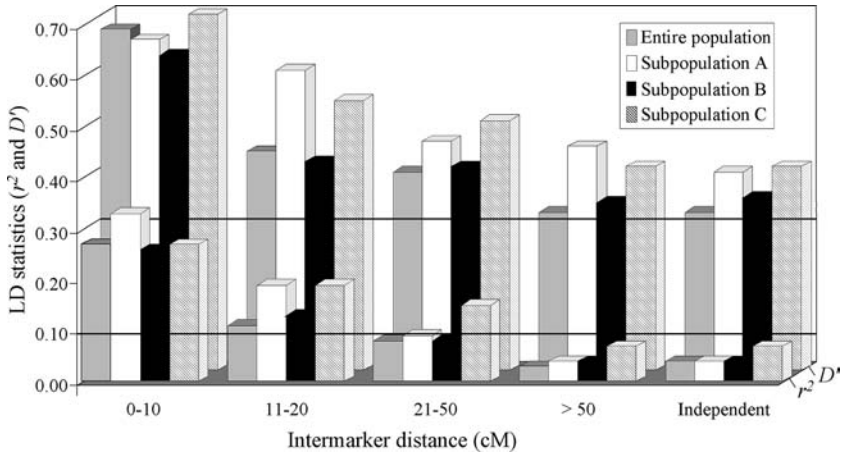


Figure 5. Average long-range LD between SSR loci. Pair-wise SSR markers have been classified based on the intermarker genetic distance (Roder et al. 1998) as follows: tightly to moderately linked (0–10 and 11–20 cM apart), loosely linked (21–50 cM), with a distance > 50 cM and independent. LD analysis was performed on the entire population (134 accessions) and on each of the three main subpopulations separated according to the STRUCTURE results with $K = 3$. Subpopulations A, B and C comprise 54 North American related accessions, 48 recent CIMMYT-ICARDA derived accessions and 32 old Mediterranean or Italian accessions, respectively. For each class, the average r^2 and D' values have been reported in front and back, respectively. Lines corresponding to $D' = 0.30$ and to $r^2 = 0.10$ have been reported.

erately linked locus pairs (intermarker distance < 20 cM) had lower frequencies.7

The percentage of locus pairs found in LD with p value < 0.01 (multifactorial permutation analysis) is reported in Figure 4 and the D' and r^2 LD measures are summarized in Figure 5.

Considering the entire population, mean p and D' values (Figures 4 and 5) indicated the presence of a very high level of LD between markers located within the same chromosome as well as on different chromosomes: the majority of tightly to moderately linked marker pairs (≤10 and 11–20 cM apart) were found in significant LD and their D' values averaged 0.67 and 0.43, respectively. The LD calculated for marker pairs > 50 cM apart and for independent pairs were still remarkably high, with significant ($p < 0.01$) LD found in 52 and 54% of the SSR pairs, respectively, and with an average D' value equal to 0.31 in both classes.

The effect of population structure on LD was accounted for by re-evaluating LD statistics within each of the three subpopulations identified by structure analysis with $K = 3$. As compared to the results obtained with the entire set of accessions, the frequency of locus pairs showing significant ($p < 0.01$) LD within each subpopulation, while remaining high for tightly to moderately linked locus pairs (< 20 cM), was greatly reduced when considering the independent locus pairs (Figure 4). A further reduction of the proportion of independent locus pairs with significant ($p < 0.01$) LD was obtained considering the subpopulations identified by the structure analysis with $K = 4$ (data not reported); in this case, the highest level of LD was observed within the North American subpopulation.

The scatter plots of the LD values as a function of the intermarker distance for the entire collection and for the three considered subpopulations are reported in Figure 6. As to the p values of the entire collection, a high variability was observed for all the intermarker distance classes; when considering the subpopulations separately the variability decreased at increasing intermarker distances. No appreciable differences in the D' ands r^2 variability were found among the entire set of accessions and the three subpopulations.

In all cases (entire collection and three subpopulations), the variation of LD statistics shown by independent locus pairs was similar (data not reported) to that shown by the locus pairs with intermarker distances > 50 cM.

Correlations between LD statistics (i.e., $-\log_{10}$ of p value, D' and r^2) concerning the intrachromosomal SSR pairs and the genetic distances were negative and, even though moderately low, significant (data not reported). The r values computed considering all 134 accessions were lower than those obtained considering separately the subpopulations; as an example, the correlation of $-\log_{10}$ p value vs. genetic distance was -0.20 in the whole population, and ranged from -0.34 to -0.45 when computed within subpopulations.

Discussion

Genetic diversity in the elite durum wheat gene pools

The mean indices of polymorphism detected by dinucleotide SSRs in the 93 cvs., selected to represent the genetic diversity present in the elite germplasm currently cultivated in the main durum wheat producing areas, confirm the findings reported by Maccaferri et al. (2003) and are similar to those reported in elite bread wheat and barley genotypes (Melchinger et al. 1994; Manifesto et al. 2001; Huang et al. 2002).

It is interesting to note that the mean diversity index of SSRs in the selected subset remained almost unchanged as compared to the complete collection herein tested, which includes, in addition, a number of either old or not widely distributed accessions. This result indicates that no noticeable erosion of the genetic basis of the elite durum wheat germplasm has occurred in the recent past. Additionally, the main gene pools are characterized by different levels of diversity and it should be noted that intensive breeding efforts specifically aimed at broadening the genetic bases of the elite germplasm are underway, particularly in the CIMMYT and ICARDA breeding programs (Autrique et al. 1996; Pfeiffer et al. 2000; Nachit and Elouafi 2004). A low level of genetic diversity characterized the modern gene pool bred in the Great Plains of the northern US (North Dakota) and Canada (in agreement with their pedigree), a finding which could in part be justified by the relatively low environmental variability present in these regions and in part by the high selection pressure applied for grain quality and uniformity.

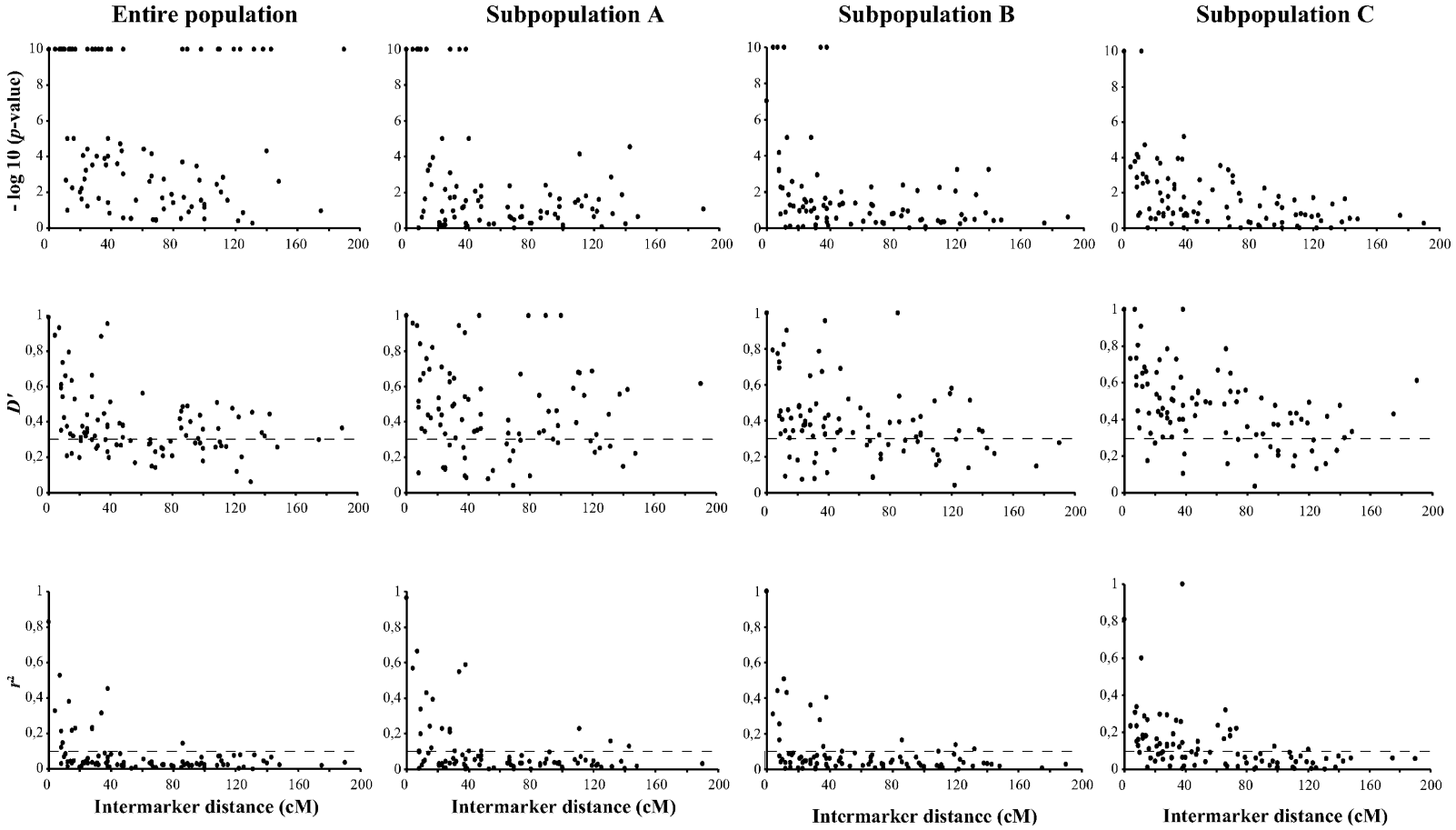


Figure 6. Scatterplot of the LD significance ($\log_{10}$ *p*-values) and statistics (D' and r^2) of marker pairs as a function of the intermarker distance (cM). LD analysis was performed on the entire populations (134 accessions) and on each of the three main subpopulations separated according to the STRUCTURE results with $K = 3$. Subpopulations A, B and C comprise 54 North American related accessions, 48 recent CIMMYT-ICARDA derived accessions and 32 old Mediterranean or Italian accessions, respectively. Lines corresponding to $D' = 0.30$ and to $r^2 = 0.10$ have been reported.

Recent efforts have aimed at incorporating novel genetic diversity in this gene pool (Clarke et al. 2001a,b). Unlike this group of materials, cvs. recently released in France showed various degrees of relatedness with the most relevant and historical gene pools and breeding lineages world-wide, thus underlying the different requirements in terms of adaptation to the wide range of environments present in the centre and south of France. It is also interesting to notice that the elite modern Italian gene pool, while being closely linked to the CIMMYT germplasm, showed evidence of genetic uniqueness, which traces mainly to old, important founders such as Cappelli and Capeiti 8.

The AMOVA as well as the distance- and model-based clustering analyses indicated that the germplasm herein analysed is highly structured. In fact, although the AMOVA evidenced significant differences among cvs. grouped on the basis of their origin, a high degree of variability was also detected within each group.

Genetic relationships and population structure

Cluster analysis showed a complex pattern of genetic relationships among the 134 accessions, structured in a number of distinct breeding lineages, in close agreement with pedigree data and often unrelated to the geographical origin of materials. Cluster analysis using a SAHN method showed in detail the genetic relationships among accessions at high levels of similarity. However, distance-based methods often introduce distortions and simplifications in the representation of the relationships among members of a large cluster (Sneath and Sokal 1973) and/or when accessions are related to two or more distinct clusters, as in the case of many modern cvs. obtained by crossing parents from different gene pools. For these reasons, accessions were also grouped using a model-based clustering method (Pritchard et al. 2000). The model applied in our study (admixture linkage model with correlated allele frequencies; Falush et al. 2003) allowed for better representing the relationships among the main gene pools present in the collection and for clarifying the mixed ancestries of several accessions and even breeding groups, such as the 'desert durums'.

LD level in durum wheat

Association results of SSR pairs (especially those less than 20 cM apart) pointed out the presence of an extensive LD, both in the entire germplasm set and within each of the main subgroups, as expected from the medium to high level of co-ancestries among genotypes. Similar levels of long-range LD, extended over several cM, have been reported for self-pollinating species such as *Arabidopsis*, when considering isolated populations (Nordborg et al. 2002), barley (Russell et al. 2003) and soybean (Hyten et al. 2004).

Population structure produced strong effects on the overall LD, considerably inflating its value: in fact, considering the entire durum wheat collection, the proportion of independent loci showing significant LD was noticeably high. In maize, an outcrossing species for which LD data have already been published, Remington et al. (2001) calculated LD among 47 SSR loci distributed across the maize genome and found considerably lower levels of LD (9.7% of SSR pairs showing LD at $p < 0.01$ in a collection of 102 lines) than in our study on durum wheat. However, recent results from a SSR survey of elite maize germplasm, including closely related inbred lines (Liu et al. 2003), showed high levels of LD (66% considering the whole germplasm and ranging from 19 to 30% within subgroups); these data were obtained through the analysis of all the locus pairs without any distinction between linked and independent loci; most of the observed LD was attributed to residual population structure within subgroups.

In our durum wheat LD analysis, we have taken into account the presence of population structure by subdividing the germplasm into three main subpopulations; this allowed us to avoid the occurrence of sample excessively small in size notwithstanding that the structure analysis actually pointed out the presence of further significant germplasm stratification within the collection. A substantial reduction of the spurious associations due to population structure has been obtained; in fact, the percentage of loosely linked and independent locus-pairs in significant LD highly decreased after accounting for the presence of three main subpopulations (Figure 4).

Since each subpopulation comprises a relatively small number of accessions, it can be argued that the observed drop in LD significance could be

rather due to a reduced power of the LD test (see Remington et al. 2001; Liu et al. 2003). However, the reduction in the number of marker pairs in significant LD (Figure 4) was noticeable for intermarker distances higher than 21 cM, while being quite low for tightly to moderately linked locus-pairs, thus meaning that the LD which characterize these locus-pairs is to be mainly attributed to linkage. Such findings are less evident when considering LD values (Figure 5), instead of LD significances (Figure 4). It should be noted that, in this study, D' values are overall high, across populations and intermarker distances (Figure 6), and frequently higher than the generally accepted D' threshold for LD (i.e. $D' = 0.3$). These results could be ascribed to the relatively small number of accessions in the subpopulations. In this respect, Mohlke et al. (2001) evidenced that the D' statistic is highly affected and inflated by sample size and allele frequencies. Moreover, it is expected an increase of the average level of relationships among accessions within subpopulation; this in turn can affect the D' values. A low level of variation, especially at high intermarker distances, has been noted for r^2 statistics in the entire collection as well as in all the subpopulations.

Highly informative SSR loci such as those used in this study (i.e., dinucleotides, rich in total number of alleles as well as in rare and subpopulation-specific alleles) are particularly valuable for evaluating the presence of population structure (Pritchard and Rosenberg 1999; Rosenberg et al. 2001). As to the detection of long-range LD, microsatellites with three or more alleles and high information content can be favourably compared to SNPs, as demonstrated by studies conducted on the human species (reviewed in Varilo and Peltonen 2004). However, the use of SSR markers also complicates the LD analysis, especially if sample size is not adequately large.

Conclusive remarks

We underline the importance of establishing reference collections including important foundation genotypes and hallmark cvs. characterized by significant long-range LD extending on a cM-wide scale (as in our study), hence suitable for gene/QTL discovery. However, the presence of a high level of structure within the elite durum wheat germplasm evidenced in our study should be duly considered in future association mapping studies. Reference collections would also permit the optimization of studies aimed at assigning new accessions to populations and at identifying the most suitable molecular platforms (e.g. sets of SSRs, SNPs, etc.) to address population structure and ancestry estimates. Reference collections will also be valuable for organizing in a unique, integrated framework the genetic and phenotypic data generated by different studies, as advised by Rosenberg et al. (2001).

Future breeding practices in durum wheat and other crops will increasingly benefit from association mapping studies. Harnessing of agronomically valuable allelic diversity will be facilitated through approaches based on association mapping with reference panels including highly diversified accessions such as the one herein assembled. Although the high LD level of this reference durum wheat collection will only allow for a coarse mapping of the genetic factors affecting the variability of the target traits, this information coupled with the analysis of haplotype profiles of segregating populations/breeding lines will offer novel opportunities for marker-assisted selection (Morgante and Salamini 2003; Peleman and van der Voort 2003). Additional reference collections characterized by low LD should also be developed to provide a level of genetic resolution sufficiently high to more precisely identify the gene/QTLs affecting the traits of interest and/or to validate the role of candidate genes. Finally, genomic platforms such as high throughput discovery of allelic variants (by ECOTILLING survey; Henikoff and Comai 2003) and fine haplotype characterization (by using mapped ESTs and SNPs; Somers et al. 2003) can be valuably applied to reference collections to assess the molecular variability present in the germplasm.

Acknowledgements

We thank Mike Ambrose (John Innes Centre), Françoise Blouet (GEVES), John Clarke (Agriculture and Agri-Food Canada), Enzo De Ambrogio (Società Produttori Sementi), Elias Elias (North Dakota State University), Franziska Löschenberger (Probstdorfer Saatzucht), Giuseppe Merisio (ENSE, Milan), Miloudi Nachit (ICARDA), Stefano Ravaglia (Società Italiana Sementi), Conxita Royo (IRTA), Kim Shantz (Western

Plant Breeder) and Rex Thompson (World Wide Wheat) for providing pure seeds and pedigree data of the cultivars herein analysed and Silvio Salvi for helpful discussions about genetic diversity analysis and data interpretation. This research was supported by the University of Bologna, Italy. Contribution of the Interdepartmental Centre for Biotechnology, University of Bologna.

References

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

Barrett B.A. and Kidwell K.K. 1998. AFLP-based genetic diversity assessment among wheat cultivars from the Pacific Northwest. Crop Sci. 38: 1261–1271.

Buckler E.S. IV and Thornsberry J.M. 2002. Plant molecular diversity and applications to genomics. Curr. Opin. Plant Biol. 5: 107–111.

Clarke J.M., Mcleod J.G., DePauw R.M., Marchylo B.A., McCaig T.N., Knox R.E., Fernandez M.R. and Ames N. 2001a. Registration of 'AC Pathfinder' Durum Wheat. Crop Sci. 41: 269–270.

Clarke J.M., Mcleod J.G., DePauw R.M., Marchylo B.A., McCaig T.N., Knox R.E., Fernandez M.R. and Ames N. 2001b. Registration of 'AC Navigator' Durum Wheat. Crop Sci. 41: 270–271.

Devos K., Penner G. and Lee S. 1995. Characterization of loci containing microsatellite sequences among Canadian wheat cultivars. Genome 38: 1037–1040.

Estoup A., Jarne P. and Cornuet J.M. 2002. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. Mol. Ecol. 11: 1591–1604.

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

Excoffier L., Smouse P. and Quattro J. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 136: 343–359.

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

Farnir F., Coppieters W., Arranz J.J., Berzi P., Cambisano N., Grisart B., Karim L., Marcq F., Moreau L., Mni M., Nezer C., Simon P., Vanmanshoven P., Wagenaar D. and Georges M. 2000. Extensive genome-wide linkage disequilibrium in cattle. Genome Res. 10: 220–227.

Flint-Garcia S.A., Thornsberry J.M. and Buckler E.S. IV. 2003. Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. 54: 357–374.

Garris A.J., McCouch S.R. and Kresovich S. 2003. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). Genetics 165: 759–769.

Gower J.C. 1972. Measures of taxonomic distance and their analysis. In: Weiner J.S. and Huizinga J. (eds), The assessment of population affinities in man. Clarendon Press, Oxford, pp. 1–24.

Hedrick P.W. 1987. Gametic disequilibrium measures: proceed with caution. Genetics 117: 331–341.

Henikoff S. and Comai L. 2003. Single-nucleotide mutations for plant functional genomics. Annu. Rev. Plant Biol. 54: 375–401.

Huang Q., Borner A., Röder S. and Ganal W. 2002. Assessing genetic diversity of wheat (*Triticum aestivum* L.) germplasm using microsatellite markers. Theor. Appl. Genet. 105: 699–707.

Hyten D.L., Song Q. and Cregan P.B. 2004. Linkage disequilibrium in four soybean populations. In: Final Abstract Guide of the International Conference on Plant & Animal Genomes XII, January 10–14, San Diego, CA 204 pp.

Jannink J., Bink M.C. and Jansen R.C. 2001. Using complex plant pedigrees to map valuable genes. Trends Plant Sci. 6: 337–342.

Jansen R.C., Jannink J. and Beavis W.D. 2003. Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. Crop Sci. 43: 829–834.

Liu K., Goodman M., Muse S., Smith J.S., Buckler E. and Doebley J. 2003. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. Genetics 165: 2117–2128.

Lu H. and Bernardo R. 2001. Molecular marker diversity among current and historical maize inbreds. Theor. Appl. Genet. 103: 613–617.

Maccaferri M., Sanguineti M.C., Donini P. and 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.

Manifesto M.M., Schlatter A.R., Hopp H.E., Suarez E.Y. and Dubcovsky J. 2001. Quantitative evaluation of genetic erosion in wheat using molecular markers. Crop Sci. 41: 682–690.

Matsuoka Y., Mitchell S.E., Kresovich S., Goodman M. and Doebley J. 2002a. Microsatellites in Zea – variability, patterns of mutations, and use for evolutionary studies. Theor. Appl. Genet. 104: 436–450.

Matsuoka Y., Vigouroux Y., Goodman M.M., Goodman M.M., Sanchez J., Buckler E. and Doebley J. 2002b. A single domestication for maize shown by multilocus microsatellite genotyping. Proc. Natl. Acad. Sci. USA 99: 6080–6084.

McRae A.F., McEwan J.C., Dodds K.G., Wilson T., Crawford A.M. and Slate J. 2002. Linkage disequilibrium in domestic sheep. Genetics 160: 1113–1122.

Melchinger A.E., Graner A., Singh M. and Messmer M. 1994. Relationships among European barley germplasm: I. Genetic diversity among winter and spring cultivars revealed by RFLPs. Crop Sci. 34: 1191–1199.

Mohlke K.L., Lange E.M., Valle T.T., Ghosh S., Magnuson V.L., Silander K., Watanabe R.M., Chines P.S., Bergman R.N., Tuomilehto J., Collins F.S. and Boehnke M. 2001. Linkage disequilibrium between microsatellite markers extends beyond 1 cM on chromosome 20 in Finns. Genome Res. 11: 1221–1226.

Morgante M. and Salamini F. 2003. From plant genomics to breeding practice. Curr. Opin. Biotechnol. 14: 214–219.

Nachit M.M., Elouafi I., Pagnotta M.A., El Saleh A., Iacono E., Labhilili M., Asbati A., Azrak M., Hazzam H., Benscher D., Khairallah M., Ribaut J.M., Tanzarella O.A., Porceddu E. and Sorrells M.E. 2001. Molecular linkage map for an intraspecific recombinant inbred population of durum wheat (*Triticum turgidum* L. var. durum). Theor. Appl. Genet. 102: 177–186.

Nachit M.M. and Elouafi I. 2004. Durum wheat adaptation in the Mediterranean dryland: breeding, stress physiology, and molecular markers. In: Challenges and Strategies for Dryland Agriculture. Crop Science Society of America and American Society of Agronomy, 677 S. Segoe Rd., Madison, WI 53711, USA. CSSA special Publication no. 32.

Nordborg M., Borevitz J.O., Bergelson J., Berry C.C., Chory J., Hagenblad J., Kreitman M., Maloof J.N., Noyes T., Oefner P.J., Stahl E.A. and Weigel D. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. Nat. Genet. 30: 190–193.

Pecetti L. and Annicchiarico P. 1998. Agronomic value and plant type of Italian durum wheat cultivars from different eras of breeding. Euphytica 99: 9–15.

Peleman J.D. and van der Voort J.R. 2003. Breeding by design. Trends Plant Sci. 8: 330–334.

Peng J., Korol A.B., Fahima T., Röder M., Ronin Y.I., Li Y.C. and Nevo E. 2000. Molecular genetic maps in wild emmer wheat, *Triticum dicoccoides*: genome-wide coverage, massive negative interference, and putative quasi-linkage. Genome Res. 10: 1509–1531.

Pfeiffer W.H., Sayre K.D. and Reynolds M.P. 2000. Enhancing genetic grain yield potential and yield stability in durum wheat. In: Royo C., Di Fonzo N., Nachit M.M. and Araus J.L. (eds), Proceedings of the Seminar Durum Wheat Improvement the Mediterranean Region: New Challenges, Zaragoza, Spain, April 12–14, Options Méditerranéennes 40: 83–94.

Powell W., Machray G.C. and Provan J. 1996. Polymorphism revealed by simple sequence repeats. Trends Plant Sci. 1: 215–222.

Pritchard J.K. and Przeworski M. 2001. Linkage disequilibrium in humans: models and data. Am. J. Hum. Genet. 69: 1–14.

Pritchard J.K. and Rosenberg N.A. 1999. Use of unlinked genetic markers to detect population stratification in association studies. Am. J. Hum. Genet. 65: 220–228.

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

Rafalski A. and Morgante M. 2004. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. Trends Genet. 20: 103–111.

Remington D.L., Thornsberry J.M., Matsuoka Y., Wilson L.M., Whitt S.R., Doebley J., Kresovich S., Goodman M.M. and Buckler E.S. IV. 2001. Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc. Natl. Acad. Sci. USA 98: 11479–11484.

Reynolds J., Weir B.S. and Cockerham C.C. 1983. Estimation for the coancestry coefficient: basis for a short-term genetic distance. Genetics 105: 767–779.

Röder M.S., Korzun V., Wendehake K., Plaschke J., Tixier M.H., Leroy P. and Ganal M.W. 1998. A microsatellite map of wheat. Genetics 149: 2007–2023.

Rohlf F.J. 1997. NTSYS-pc version 2.0. Exeter Press, East Setauket, NY, USA.

Rosenberg N.A., Burke T., Elo K., Feldman M.W., Freidlin P.J., Groenen M.A., Hillel J., Maki-Tanila A., Tixier-Boichard M., Vignal A., Wimmers K. and Weigend S. 2001. Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. Genetics 159: 699–713.

Russell J.R., Booth A., Fuller J.D., Baum M., Ceccarelli S., Grando S. and Powell W. 2003. Patterns of polymorphism detected in the chloroplast and nuclear genomes of barley landraces sampled from Syria and Jordan. Theor. Appl. Genet. 107: 413–421.

Saghai-Maroof M.A., Soliman K.M., Jorgensen R.A. and Allard R.W. 1984. Ribosomal DNA spacer length polymorphism in barley: Mendelian inheritance, chromosomal location and population dynamics. Proc. Natl. Acad. Sci. USA 83: 1757–1761.

Schneider S., Roessli D. and Excoffier L. 2000. Arlequin ver. 2.000: a software for population genetics data analysis. Genetics and Biometry Laboratory, Department of Anthropology and Ecology, University of Geneva, Switzerland. http://anthro.unige.ch/arlequin/.

Slatkin M. 1994. Linkage disequilibrium in growing and stable populations. Genetics 137: 331–336.

Sneath P.H.A. and Sokal R.R. 1973. Numerical Taxonomy. W.H. Freeman and Co, San Francisco, CA, USA.

Somers D.J., Kirkpatrick R., Moniwa M. and Wash A. 2003. Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. Genome 46: 431–437.

Thiel T., Michalek W., Varshney R.K. and Graner A. 2003. Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). Theor. Appl. Genet. 106: 411–422.

Thornsberry J.M., Goodman M.M., Doebley J., Kresovich S., Nielsen D. and Buckler E.S. IV. 2001. Dwarf8 polymorphisms associate with variation in flowering time. Nat. Genet. 28: 286–289.

Varilo T. and Peltonen L. 2004. Isolates and their potential use in complex gene mapping efforts. Curr. Opin. Genet. Dev. 2004(14): 316–323.

Weir B.S. 1996. Genetic Data Analysis II. Methods for Discrete Population Genetic Data. Sinauer Associates, Inc. Publishers.