

A panel of elite accessions of durum wheat (*Triticum durum* Desf.) suitable for association mapping studies

Marco Maccaferri¹, Maria Corinna Sanguineti¹, Vincenzo Natoli¹, José Luis Araus Ortega², Moncef Ben Salem³, Jordi Bort², Cynda Chenenaoui³, Enzo De Ambrogio⁴, Luis Garcia del Moral⁵, Andrea De Montis⁴, Ahmed El-Ahmed⁶, Fouad Maalouf⁷, Hassan Machlab⁷, Mark Moragues⁸, Jihan Motawaj⁹, Miloudi Nachit⁹, Nasserlehaq Nserallah¹⁰, Hassan Ouabbou¹⁰, Conxita Royo⁸ and Roberto Tuberosa^{1*}

¹Department of Agroenvironmental Science and Technology, University of Bologna, Bologna, Italy, ²Departament de Biologia Vegetal, Universitat de Barcelona, Barcelona, Spain, ³Tunisian National Institute of Agronomic Research, Tunis, Tunisia, ⁴Società Produttori Sementi Bologna, Divisione Ricerca, Argelato (BO), Italy, ⁵Dpto. Fisiología Vegetal, University of Granada, Granada, Spain, ⁶Plant Protection Department, Aleppo University, Aleppo, Syria, ⁷Department of Plant Breeding, Lebanese Agricultural Research Institute, Bekaa, Lebanon, ⁸Area de Conreus Extensius Centre UdL-IRTA, Lleida, Spain, ⁹ICARDA, Aleppo, Syria and ¹⁰CRRA-INRA, Settat, Morocco

Received 15 February 2006; Accepted 17 February 2006

Abstract

The effectiveness of association mapping (AM) based on linkage disequilibrium (LD) is currently being tested in a number of crops. An important prerequisite for the application of AM is the availability of collections of accessions with a suitable level of genetic variation for target traits and with limited spurious LD due to the presence of population structure. Herein, the results of a genome-wide molecular characterization of a collection of elite durum wheat accessions well-adapted to Mediterranean environments are presented. Ninety-seven highly polymorphic simple sequence repeats and 166 amplified fragment length polymorphism markers were used to characterize 189 durum accessions, mainly cultivars and advanced breeding lines. Genome-wide significant and sizeable LD indices at a centimorgan scale were observed, while LD mainly decayed within 10 cM. On the other hand, effects due to spurious LD were notably lower than those previously observed in a durum wheat collection sampling durum gene pools of more diverse origin. These results, coupled with the high level of genetic variability detected for a number of important morpho-physiological traits and their high heritability, indicate the suitability of this collection for AM studies targeting agronomically important traits.

Keywords: AFLP; association mapping; durum wheat; linkage disequilibrium; SSR; *Triticum durum*

Introduction

The introduction of genomics approaches has greatly accelerated the identification of chromosome regions

harbouring genes/quantitative trait loci (QTL) controlling agronomic traits in crops (Morgante and Salamini, 2003; Varshney *et al.*, 2005). Among such approaches, association mapping (AM) is receiving increasing attention as an additional means to harness beneficial genetic variation present in germplasm collections (Buckler and Thornsberry, 2002; Flint-Garcia *et al.*, 2003; Jansen *et al.*, 2003; Rafalski and Morgante, 2004; Breseghello and

* Corresponding author: Department of Agroenvironmental Sciences and Technology, Via Fanin, 44, 40127-Bologna, Italy.
E-mail: roberto.tuberosa@unibo.it

Sorrells, 2006). In simple terms, the primary objective of AM studies is to detect correlations between genotypes and phenotypes in a sample of individuals, based on linkage disequilibrium (LD; Zondervan and Cardon, 2004). LD is the non-random combination of alleles at two genetic loci, which in random-mating populations is mostly generated by mutation and genetic drift, and is subsequently dissipated by recombination (Pritchard and Przeworski, 2001; Varilo *et al.*, 2003).

In contrast with QTL analysis, which is based on biparental mapping populations, AM is characterized by a higher probability of type I error (false positive) caused by unaccounted subdivisions in the sample (i.e. population structure; Pritchard *et al.*, 2000a, 2000b) which can bias the estimates of allele effects. Additionally, the type II (lack of statistical power) error is increased, due to (i) lower correlation between markers and genes caused by LD decay, (ii) unbalanced design due to alleles present at different frequencies and (iii) a multiple-testing problem which results in high genome-wide significance thresholds (Gordon and Finch, 2005). For these reasons, it has been suggested that AM may have limited application for the detection of rare variants at loci which are variable between populations, but are prevalently fixed within subpopulations (Brescghello and Sorrells, 2006; Flint-Garcia *et al.*, 2005). For a genome-wide search, collections of materials characterized by a high LD (slow LD decay rate as a function of genetic distance) are preferable due to the reasonably low number of markers required to reveal a significant marker–trait association. Moreover, based on the extent of LD present among the materials, it is possible to identify conserved haplotypes, evaluate their phenotypic effects and readily enter them in marker-assisted selection programmes.

In durum wheat (*Triticum durum* Desf.), a selfing species, the high level of LD revealed in elite accessions representing the main breeding groups suggests their suitability for QTL discovery, as demonstrated in a recent study targeting resistance to soil-borne wheat cereal virus (Maccaferri *et al.*, 2005a). Based on these encouraging results, in order to apply AM to identify QTL influencing morpho-physiological traits involved in the adaptive response of durum wheat to drought, a second set of elite accessions has been established within an EU-funded project (IDuWUE; for more details see the website, <http://www.distagenomics.unibo.it/iduwue/index.html>) aimed at improving water-use efficiency of this important Mediterranean crop. We report here the results pertaining to the genetic structure of this new set of accessions and the extent of LD throughout the wheat genome as assessed with simple sequence repeat (SSR) markers.

Materials and methods

Plant materials

For this study, we chose 189 durum wheat accessions, mainly cultivars and advanced lines representing a large part of the genetic diversity present in the major improved durum wheat gene pools adapted to the Mediterranean environments (Table 1 can be accessed at <http://www.distagenomics.unibo.it/iduwue/publications.html>). According to the country/institution of origin, the accessions were sampled as follows: (i) 84 accessions from the International Center for Agricultural Research in the Dry Areas (ICARDA); (ii) 41 Italian accessions, comprising some important old ‘founder accessions’, also used in the International Maize and Wheat Improvement Center (CIMMYT)/ICARDA breeding programmes; (iii) 21 accessions from Spain; (iv) 18 accessions from CIMMYT, some of which have been used in national breeding programmes throughout the Mediterranean Basin; (v) 14 accessions from Morocco and two from Tunisia; (vi) nine accessions representative of the germplasm cultivated under irrigation in the south-western region of the USA (Arizona/California) and commonly referred to as ‘desert durums’; these accessions are related to the CIMMYT and Italian materials.

Molecular data

For each accession, DNA was extracted from the leaves of 20 seedlings (Saghai-Marooif *et al.*, 1984). A set of 97 genomic SSR loci was utilized in the present study (Table 2 can be accessed at <http://www.distagenomics.unibo.it/iduwue/publications.html>). Most of the loci were chosen on the basis of the mapping results of a joint project based on a recombinant inbred line population originating from two parental cultivars included in the collection, followed by a pre-screening for polymorphism on eight founders of the durum collection. The SSRs included public BARC, CFA, CFD, WMC and WMS sets (Röder *et al.*, 1998, Somers *et al.*, 2004, Sourdille *et al.*, 2004) and a small set of restricted GWM primers (property of TraitGenetics, Gatersleben, Germany). Out of the 97 amplified and polymorphic loci, 19 could not be reliably positioned on either the International Triticeae Mapping Initiative (ITMI) map, the consensus wheat maps or on our durum wheat map, and hence these markers were not considered for the LD analysis. The SSR loci mapped to all 14 durum wheat chromosome pairs (Table 2). Additionally, 166 polymorphic amplified fragment length polymorphism (AFLP) loci originating from six primer combinations

(*Sse8387I/MseI* restriction enzyme combination) were included in the co-ancestry analysis. SSR and AFLP amplification products were obtained with minor modifications to the protocol of Röder *et al.* (1998) and of Vos *et al.* (1995), respectively. SSR and AFLP analyses were carried out using fluorescently labelled primers in the LI-COR Model 4200 IR² sequencer (LI-COR, Lincoln, NE, USA).

Field trials and phenotypic data

The 189 accessions of the elite Mediterranean durum collection were evaluated in 2004 in field trials carried out under rainfed conditions in the following environments: Cadriano (Bologna), Italy; Cerignola (Foggia), Italy; Lleida, Spain; Sidi El Aydi (Settat), Morocco; and Tel Amara, Rayack (Bekaa), Lebanon. Field trials consisted of unreplicated plots (4 m²) arranged according to a modified augmented design with three checks. Field trials in Morocco and Lebanon were carried out under both rainfed and irrigated conditions. In total, more than 30 morphophysiological traits involved in the response to drought were considered. Herein, we report only some of the results on leaf posture, heading date, growing degree days (GDD; from emergence to heading), plant height, peduncle length, mean kernel weight and chlorophyll content as measured with a Minotta SPAD 502 Chlorophyll Meter hand-held instrument (indexed SPAD units). Further details on the phenotypic data are reported at the IDuWUE website and will soon be published in a companion paper (M. Maccaferri *et al.*, unpublished results).

Kinship and LD analysis

Pairwise genetic similarity values (GS_{ij}) among all accessions were calculated (NTSYS-pc software version 2.0; Rohlf, 1997) as the proportion of loci with shared alleles (Lu and Bernardo, 2001; Matsuoka *et al.*, 2002), by using the simple matching coefficient for multi-state qualitative data. Both SSR and AFLP data were considered. A 189 × 189 matrix of co-ancestry (kinship) coefficients among accessions, calculated according to Loiselle *et al.* (1995) as correlation coefficients between allelic states, was obtained from the molecular data using the program SPAGeDi (Hardy and Vekemans, 2002). The gross genetic diversity structure of the collection herein analysed was investigated with a model-based (Bayesian) clustering algorithm (STRUCTURE software; Pritchard *et al.*, 2000a, 2000b; Falush *et al.*, 2004), which identifies subgroups of accessions with distinct allele frequencies. Distinct from the cluster analysis, in this model-based method each accession is permitted membership of several different subgroups, with membership coefficients

totalling 1. The program was run for a number (K value) of hypothetical subpopulations ranging from two to eight. Runs were carried out by setting 100,000 iterations, of which only the last 50,000 were recorded, and assuming an admixture linkage model with correlated allele frequencies. Mapped SSR loci were used to calculate whole-genome LD estimates (r^2 , D' and P -values) for each pair of markers; weighted averages of r^2 and D' for loci with multiple alleles were estimated (Farnir *et al.*, 2000). LD P -values were calculated as the proportion of permuted gamete distributions less probable than the observed gamete distribution under the null hypothesis of independence (Weir, 1996), with 10,000 permutations. Calculations were carried out in TASSEL 1.9.3. (www.maizegenetics.net; Yu *et al.*, 2006).

Results and discussion

Genetic diversity structure and LD levels

In the model-based analysis carried out with STRUCTURE software, the admixture linkage model with correlated allele frequencies was chosen to account for the complex genetic structure of the collection, i.e. for the various levels of inter-relationships among individuals, and to better quantify such relationships. In fact, the pedigree data of several genotypes indicated a mixed origin due to crosses between ancestors belonging to different gene pools. The main parameter, i.e. the Bayesian posterior probability of data, linearly declined by using a K number of subpopulations from two to eight, with the alpha parameter (Dirichlet parameter for degree of admixture) dropping from 0.56 for K = 3 to 0.25 for K = 6 and to 0.16 for K = 8, suggesting that at least six to eight subgroups could be consistently identified, where a global alpha parameter value less than 0.20 is considered as a good indication of consistent assignment of an entry to specific subpopulations. Nevertheless, even when considering K values of six and eight, a consistent number of accessions showed a high degree of admixture. With K = 8, only 111 accessions out of the 189 showed a membership coefficient to one specific subpopulation higher than 0.50 (i.e. 50% or more of the genome inherited from the ancestors of one specific subpopulation).

For each accession, membership coefficients to the eight subpopulations are reported in Table 1. The distribution of accessions among subpopulations was as follows: subpopulation 1 with ICARDA and other accessions originated from Syrian materials (e.g. derived from the Omrabi-group); subpopulation 2 with CIMMYT, ICARDA, Italian, Spanish and Moroccan cultivars all related to the hallmark CIMMYT founder Altar

84 (released as Gallareta in Spain), widely used in the breeding programmes carried out in the 1980s and 1990s; subpopulation 3 with a group of ICARDA (e.g. Cham-1, Stojocri-3, Loukos-1, etc.) and Italian (e.g. Creso, Produra, etc.) materials related to the original CIMMYT releases; subpopulations 4, 5 and 6 with subgroups of ICARDA materials; subpopulation 7 with Italian accessions obtained from Cappelli and Valnova founders; subpopulation 8 with a breeding lineage originated from the successful CIMMYT cross Jo/Aa/Fg widely used in several Mediterranean countries to derive successful cultivars, such as Duilio, Karim and Yasmine. These findings were also confirmed by cluster analysis (data not shown) which did not indicate the presence of highly differentiated subpopulations: all main nodes of the cluster branched within a narrow genetic similarity interval (from 0.55 to 0.65). Only a limited group of about 20 Italian accessions mostly derived from local materials could be differentiated from the large majority of materials related to the CIMMYT/ICARDA gene pool. Rather, a number of subpopulations, corresponding to different breeding lineages, could be distinguished at higher similarity levels in the collection; these subpopulations often included fewer than 15 accessions each, thus indicating the presence of a complex pattern of familial relationships among accessions. Within breeding lineage, various levels of relatedness were noted, as evidenced by the wide range of similarity levels at which the subcluster nodes branched within lineages (data not shown).

The results of both model-based and distance-based (cluster) analysis revealed a genetic structure of this collection rather different from that previously observed in the set of elite durum accessions investigated by Maccaferri *et al.* (2005b) who demonstrated the presence of a highly structured diversity, with major differences between the North American and the Mediterranean entries. Our data also highlighted the accessions with complex genetic relationships (across breeding lineages), in a higher frequency as compared to that observed among the durum materials tested by Maccaferri *et al.* (2005b). It should be underlined that the elite accessions evaluated here mainly derive from crosses among CIMMYT/ICARDA and local materials that have been/are being released in Mediterranean countries, while the previous durum wheat collection included many accessions from the durum gene pool selected in the Great Plains of the USA and Canada (North American group) as well as French materials poorly adapted to the Mediterranean environments because of day-length sensitivity and late flowering (Maccaferri *et al.*, 2005b). Only 42 accessions, mainly Italian cultivars, are common to both collections.

The calculation of the co-ancestry (kinship) coefficients among accessions (Loiselle *et al.*, 1995) allowed

us to compare the average level of relationships among the accessions to the average level of co-ancestry present among the set of 277 maize inbred lines assembled for LD analysis (Flint-Garcia *et al.*, 2005) (Table 3). Interestingly, the distribution of the co-ancestry coefficients between these two collections is markedly similar; this finding illustrates that the presence of a residual subtle structure due to familial relationships is a common feature of the elite germplasm of crop species.

Decay of linkage disequilibrium

The use of mapped SSR markers allowed us to estimate the average degree of LD present at the genome-wide level in the elite durum collection; similar investigations have been carried out also in *Arabidopsis* (Nordborg *et al.*, 2002), maize (Remington *et al.*, 2001) and durum wheat (Maccaferri *et al.*, 2005b). The 3003 possible pairs of mapped SSRs were subdivided into five groups: 26 tightly linked pairs (inter-marker distances ≤ 10 cM), 14 linked pairs (from > 10 to ≤ 20 cM), 35 loosely linked pairs (from > 20 to ≤ 50 cM), independent but syntenic pairs (from > 50 to < 300 cM) and independent pairs (different chromosomes). The results are reported as scatterplots in Fig. 1. Using the widely accepted LD thresholds (i.e. 0.1 for r^2 and 0.3 for D'), two main findings emerge: (i) noticeable LD extending on a cM scale among the tightly linked (≤ 10 cM) marker pairs; (ii) a rather low level of spurious associations due to population structure in the elite Mediterranean collection as compared to the highly

Table 3. Distribution of co-ancestry coefficients of the durum wheat collection herein described (189 accessions) as compared to that of a maize collection suitable for LD and association mapping studies

| Co-ancestry coefficient class | Frequency | |
|-------------------------------|-------------|-------|
| | Durum wheat | Maize |
| 0 | 0.549 | 0.561 |
| > 0 to ≤ 0.1 | 0.289 | 0.236 |
| > 0.1 to ≤ 0.2 | 0.104 | 0.122 |
| > 0.2 to ≤ 0.3 | 0.033 | 0.047 |
| > 0.3 to ≤ 0.4 | 0.014 | 0.015 |
| > 0.4 to ≤ 0.5 | 0.006 | 0.005 |
| > 0.5 to ≤ 0.6 | 0.003 | 0.003 |
| > 0.6 to ≤ 0.7 | 0.002 | 0.002 |
| > 0.7 to ≤ 0.8 | 0 | 0.001 |
| > 0.8 to ≤ 0.9 | 0 | 0.002 |
| > 0.9 to ≤ 1.0 | 0 | 0.006 |

Data of the maize collection (277 lines assembled by Flint-Garcia *et al.*, 2005) were obtained from the website <http://www.maizegenetics.net/>. Co-ancestry coefficients of the durum collection have been obtained from 97 SSR and 166 AFLP markers.

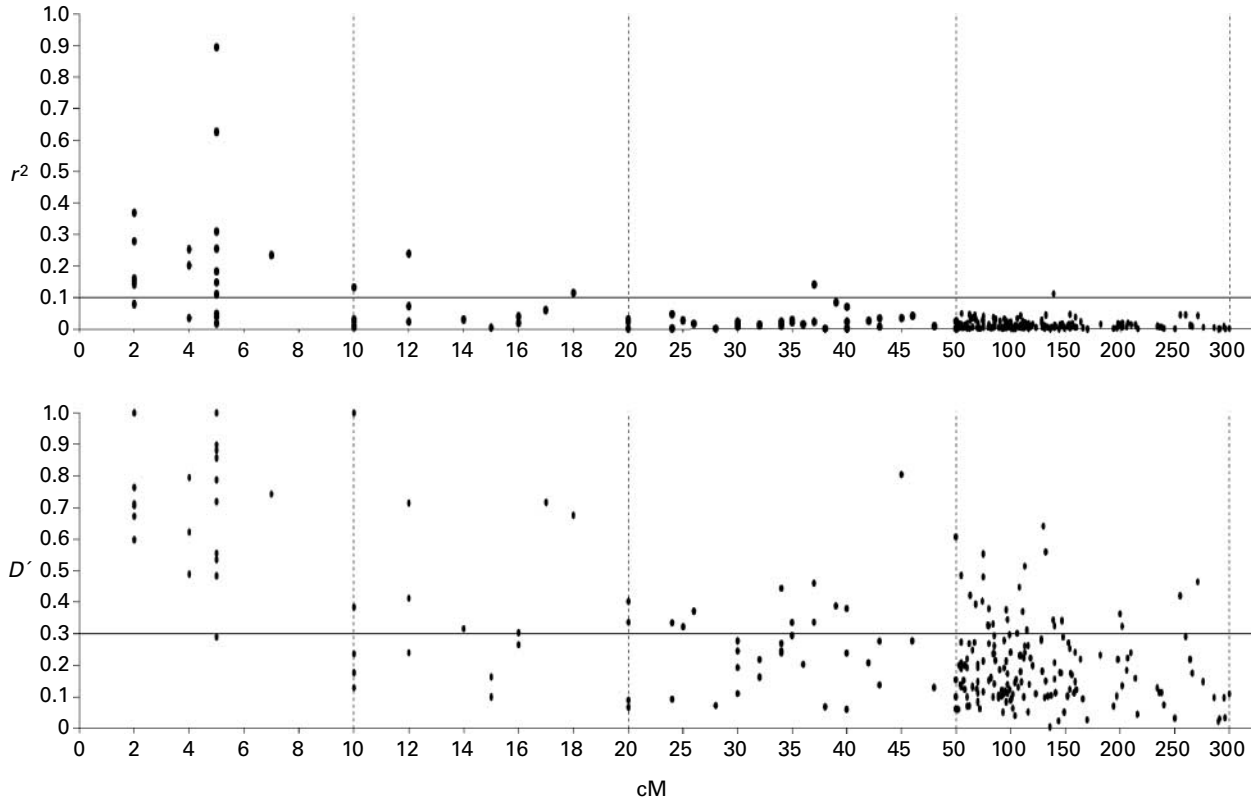


Fig. 1. Scatterplots of LD values (r^2 and D') for intra-chromosomal loci pairs, as a function of the inter-marker distance (cM); data from 235 SSR marker pairs.

stratified collection obtained from worldwide elite gene pools (Maccaferri *et al.*, 2005b).

As to the LD portion that can be mainly ascribed to linkage effects, sizeable values were observed in the <10 cM class of inter-marker distances, with average r^2 and D' values equal to 0.18 and 0.62, respectively, and an average frequency of marker pairs with $P < 0.01$ (permutation test) equal to 0.81 (Table 4). Nevertheless, these pairs showed LD values lower than the general threshold

of 0.1 for r^2 and 0.3 for D' ; this result can be ascribed to the genome-wide irregularity of the LD-decay patterns analogously to what has already been observed in collections of *Arabidopsis* (Hagenblad *et al.*, 2004) and maize (Rafalski and Morgante, 2004). At increasingly higher distances between markers on the same chromosome, lower r^2 and D' values were obtained (r^2 from 0.05 to 0.01 and D' from 0.34 to 0.20, in case of linked and independent marker pairs, respectively). Table 4 reports also the LD

Table 4. Comparison of LD descriptive statistics between the Mediterranean durum wheat collection herein described (189 accessions) and the collection of durum wheat accessions of various origins characterized by Maccaferri *et al.* (2005b)

| Inter-marker distance (cM) | No. of pairs | | r^2 | | D' | | Mean P -value | | Pairs with $P \leq 0.01$ | |
|------------------------------------------|----------------|----------------|-------|--------------------|------|--------------------|-----------------|--------------------|--------------------------|------|
| | A ^a | B ^b | A | B | A | B | A | B | A | B |
| ≤ 10 | 26 | 9 | 0.18 | 0.27 ^{ns} | 0.62 | 0.67 ^{ns} | 0.03 | 0 ^{ns} | 0.81 | 1.00 |
| > 10 to ≤ 20 | 14 | 11 | 0.05 | 0.13* | 0.34 | 0.46 ^{ns} | 0.14 | 0.01 ^{ns} | 0.43 | 0.91 |
| > 20 to ≤ 50 | 35 | 30 | 0.02 | 0.07** | 0.26 | 0.38** | 0.21 | 0.02** | 0.31 | 0.78 |
| > 50 to end of chromosome ^c | 160 | 46 | 0.01 | 0.03** | 0.20 | 0.31** | 0.28 | 0.08** | 0.14 | 0.54 |
| Inter-chromosomal ^d | 2768 | 1558 | 0.02 | 0.04** | 0.25 | 0.31** | 0.25 | 0.11** | 0.23 | 0.53 |

Population A versus population B comparison: ns, not significant; *significant at $P \leq 0.05$; **significant at $P \leq 0.01$.

^aData referred to the elite Mediterranean durum wheat collection (189 accessions).

^bData referred to the elite durum wheat collection (134 accessions) previously characterized by Maccaferri *et al.* (2005b).

^cIndependent intra-chromosomal marker pairs.

^dIndependent inter-chromosomal marker pairs.

Table 5. Mean, minimum and maximum values across environments for some morpho-physiological traits recorded in the durum collection; for each trait, heritability (on a mean basis) and the percentage of phenotypic variation explained by population structure ($K = 8$) have also been reported

| Trait | No. of environments | Mean | Minimum | Maximum | h^2 (%) | R^2 (%) |
|-----------------------------------|---------------------|------|---------|---------|-----------|-----------|
| Leaf posture (S) ^a | 2 | 6 | 2 | 8 | 67.1 | 20.5 |
| Heading date (days) ^b | 7 | 128 | 121 | 135 | 88.2 | 2.6 |
| GDD (°C) ^c | 7 | 1277 | 1188 | 1369 | 89.2 | 4.1 |
| Plant height (cm) | 5 | 80 | 69 | 112 | 87.6 | 12.3 |
| Peduncle length (cm) ^d | 5 | 16 | 12 | 24 | 83.7 | 29.1 |
| Kernel weight (mg) | 7 | 44 | 39 | 50 | 85.9 | 16.6 |
| Chlorophyll content (SPAD units) | 7 | 41 | 37 | 46 | 21.6 | 8.2 |

^a Measured, at stage 49 in the Zadoks scale, as score from 1 (flag leaf completely prostrate) to 9 (completely erect).

^b Days from emergence to heading.

^c Growing degree days from emergence to heading.

^d Measured as distance from the flag leaf collar to the base of the ear.

values of different inter-marker distance classes obtained by Maccaferri *et al.* (2005b). When considering the tightly linked pair class, no significant differences (F test) were found between the r^2 and D' values of the two collections; on the contrary, the LD indices for the other inter-marker distance classes (from linked to independent pairs) were significantly lower in the elite Mediterranean collection. As with the LD decay with genetic distance and the presence of spurious LD, these results indicate that the collection is characterized by weaker effects due to population structure as compared to that described by Maccaferri *et al.* (2005b). Although these values are higher than those observed by Remington *et al.* (2001) in a collection of maize accessions, they are similar to those reported for elite collections of maize inbreds (Liu *et al.*, 2003; Flint-Garcia *et al.*, 2005). Moreover, LD patterns similar to those reported here have also been recently described in barley by Maly-sheva-Otto *et al.* (2006).

Influence of population structure on phenotypic variation of target traits

The 189 accessions were evaluated in field trials carried out during 2004 under rainfed and irrigated conditions in five Mediterranean countries. Table 5 reports their performance with respect to a number of target traits. With the exception of chlorophyll content, all traits had medium-high heritability, underlining the limited influence of environment on trait expression. Highly significant differences were detected among accessions for all traits (F test, data not shown). Only a limited proportion of the extensive variation was due to the presence of subpopulations; in fact, the percentage of variation accounted for by population structure [as measured by the determination coefficient (R^2) of the multiple regression model, including the membership coefficients of each accession to each of the eight

hypothetical subpopulations (Table 1)], though somewhat trait-dependent, did not exceed 29%. For heading date (measured either as days from emergence or as growing degree days), chlorophyll content and plant height, no appreciable effects due to population structure were observed; therefore, for these traits at least, we would expect only negligible type I and type II errors in AM. The most marked influence of population structure affected peduncle length and leaf posture; with respect to the latter, this can be readily explained by a selection effect, in particular the move to the 'erect leaf' ideotype in many breeding lineages (e.g. lines derived from Altar 84).

In conclusion, we have noted sizeable genome-wide LD at a centimorgan scale, and some limited spurious LD due to population structure. However, this latter phenomenon is less severe in the present varietal panel than in the elite durum collection analysed by Maccaferri *et al.* (2005b). These findings, coupled with the high level of genetic variability detected for a number of important morpho-physiological traits and their high heritability, are supportive of the suitability of this collection for AM studies targeting agronomically important traits.

Acknowledgements

The financial contribution of the European Union (IDuWUE project; INCO contract number ICA3-CT-2002-10 028) is gratefully acknowledged.

References

- Breseghele F and Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172: 1165–1177.
- Buckler ES 4th and Thornsberry JM (2002) Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology* 5: 107–111.

- Falush D, Stephens M and Pritchard JK (2004) Inference of population structure: extensions to linked loci and correlated allele frequencies. *Genetics* 164: 1567–1587.
- Farnir F, Coppieters W, Arranz JJ, 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 Research* 10: 220–227.
- Flint-Garcia SA, Thornsberry JM and Buckler ES 4th (2003) Structure of linkage disequilibrium in plants. *Annual Review of Plant Biology* 54: 357–374.
- Flint-Garcia SA, Thuillet AC, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM and Buckler ES (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant Journal* 44: 1054–1064.
- Gordon D and Finch SJ (2005) Factors affecting statistical power in the detection of genetic association. *Journal of Clinical Investigations* 115: 1408–1418.
- Hagenblad J, Tang C, Molitor J, Werner J, Zhao K, Zheng H, Marjoram P, Weigel D and Nordborg M (2004) Haplotype structure and phenotypic associations in the chromosomal regions surrounding two *Arabidopsis thaliana* flowering time loci. *Genetics* 168: 1627–1638.
- Hardy OJ and Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes* 2: 618–620.
- Jansen RC, Jannink JL and Beavis WD (2003) Mapping quantitative trait loci in plant breeding populations: use of parental haplotype sharing. *Crop Science* 43: 829–834.
- Liu K, Goodman M, Muse S, Smith JS, Buckler E and Doebley J (2003) Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. *Genetics* 165: 2117–2128.
- Loiselle BA, Sork VL, Nason J and Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). *American Journal of Botany* 82: 1420–1425.
- Lu H and Bernardo R (2001) Molecular marker diversity of current and historical maize inbreds. *Theoretical and Applied Genetics* 103: 613–617.
- Maccaferri M, Ratti C, Tuberosa R, Rubies Autonell C, Vallega V and Sanguineti MC (2005a) A durum wheat germplasm collection suitable for gene discovery via association mapping. In: *Proceedings of the International Symposium on Genomics-based Plant Germplasm Research*, 25–28 April 2005. Beijing, China, pp. 47–57.
- Maccaferri M, Sanguineti MC, Noli E and Tuberosa R (2005b) Population structure and long-range disequilibrium in a durum wheat elite collection. *Molecular Breeding* 15: 271–290.
- Malysheva-Otto LV, Ganai MW and Roder MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.). *BMC Genetics* 7: 6 (DOI: 10.1186/1471-2156-7-6).
- Matsuoka Y, Mitchell SE, Kresovich S, Goodman M and Doebley J (2002) Microsatellites in *Zea*—variability, patterns of mutations, and use for evolutionary studies. *Theoretical and Applied Genetics* 104: 436–450.
- Morgante M and Salamini F (2003) From plant genomics to breeding practice. *Current Opinion in Biotechnology* 14: 214–219.
- Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA and Weigel D (2002) The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics* 30: 190–193.
- Pritchard JK and Przeworski M (2001) Linkage disequilibrium in humans: models and data. *American Journal of Human Genetics* 69: 1–14.
- Pritchard JK, Stephens M and Donnelly P (2000a) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Pritchard JK, Stephens M, Rosenberg NA and Donnelly P (2000b) Association mapping in structured populations. *American Journal of Human Genetics* 67: 170–181.
- Rafalski A and Morgante M (2004) Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends in Genetics* 20: 103–111.
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM and Buckler ES 4th (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Science USA* 98: 11479–11484.
- Röder MS, Korzun V, Wendehake K, Plaschke J, Tixier MH, Leroy P and Ganai MW (1998) A microsatellite map of wheat. *Genetics* 149: 2007–2023.
- Rohlf FJ (1997) *NTSYS-PC: Numerical Taxonomy and Multivariate Analyses System*. New York: Exeter Software.
- Saghai-Marouf MA, Soliman KM, Jorgensen RA and Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proceedings of the National Academy of Science USA* 81: 8014–8018.
- Somers DJ, Isaac P and Edwards KJ (2004) A high density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). *Theoretical and Applied Genetics* 109: 1105–1114.
- Sourdille P, Singh S, Cadalen T, Brown-Guedira GL, Gay G, Qi L, Gill BS, Dufour P, Murigneux A and Bernard M (2004) Microsatellite-based deletion bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). *Functional and Integrative Genomics* 4: 12–25.
- Varilo T, Paunio T, Parker A, Perola M, Meyer J, Terwilliger JD and Peltonen L (2003) The interval of linkage disequilibrium (LD) detected with microsatellite and SNP markers in chromosomes of Finnish populations with different histories. *Human Molecular Genetics* 12: 51–59.
- Varshney RK, Graner A and Sorrells ME (2005) Genomics-assisted breeding for crop improvement. *Trends in Plant Science* 10: 561–562.
- Vos P, Hogers R, Bleeker M, Reijmans M, Van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M and Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acid Research* 23: 4407–4414.
- Weir BS (1996) *Genetic Data Analysis II*. Sunderland, MA: Sinauer Associates.
- Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S and Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics* 38: 203–208.
- Zondervan KT and Cardon LR (2004) The complex interplay among factors that influence allelic association. *Nature Reviews Genetics* 5: 89–100.