

# Association mapping in durum wheat grown in a broad range of water regimes

Maccaferri M<sup>1</sup>, Sanguineti MC<sup>1</sup>, Ben Salem M<sup>3</sup>, El-Ahmed A<sup>4</sup>, del Moral LFG<sup>5</sup>, Demontis A<sup>6</sup>, Maalouf F<sup>7</sup>, Nachit M<sup>8</sup>, Nserallah N<sup>9</sup>, Royo C<sup>10</sup>, Tuberosa R<sup>1</sup>

<sup>1</sup>*DISTA, University of Bologna, Bologna, Italy*, <sup>2</sup>*Plant Biology Dept, Barcelona, Spain*, <sup>3</sup>*INRAT, Tunis, Tunisia*, <sup>4</sup>*Plant Protection Dept, Aleppo, Syria*, <sup>5</sup>*Plant Physiology Dept, Granada, Spain*, <sup>6</sup>*Società Produttori Sementi, Argelato (BO), Italy*, <sup>7</sup>*LARI, Zahleh, Lebanon*, <sup>8</sup>*ICARDA, Aleppo, Syria*, <sup>9</sup>*CRRA-INRA, Settat, Morocco*, <sup>10</sup>*Udl-IRTA, Lleida, Spain*

## ABSTRACT

In major crops, panels of germplasm accessions have been assembled for allele mining at targeted candidate genes and for whole-genome association mapping studies. A germplasm collection of 189 cultivated durum wheat accessions representing the major breeding groups was tested in a broad range of Mediterranean environments under rainfed and irrigated conditions with an average yield ranging from 0.9 to 6.8 t/ha. Yield, yield components, agronomic and physiological traits were recorded in 15 trials and tested for significant association with the allelic profile of 180 microsatellite markers (simple sequence repeats, SSRs). Population structure was evaluated and the clustering data were used as covariates in the association test. SSRs showing significant associations in several environments were identified for the traits with the highest heritability values, i.e. plant height, heading date, peduncle length and kernel weight, with  $R^2$  values ranging from 5 to 10%. As to yield and yield components, most associations were detected only in two to four environments, with an average  $R^2$  value lower than 5%.

## INTRODUCTION

Genomics approaches have greatly accelerated the identification of chr. regions harbouring genes/QTLs (Quantitative Trait Loci) controlling agronomic traits in crops (Morgante and Salamini, 2003; Varshney and Tuberosa, 2007). Among such approaches, association mapping (AM) is receiving increasing attention as a method complementary to traditional bi-parental mapping to associate the genotype to phenotype by exploiting the genetic variation present in germplasm collections (Buckler and Thornsberry, 2002; Breseghello and Sorrells, 2006). For a genome-wide association approach, collections of materials characterized by high linkage disequilibrium (LD) levels are preferable due to the reasonably low number of markers required to reveal significant marker-trait associations. In the elite germplasm of durum wheat (*Triticum durum* Desf.), the presence of LD at a cM-wide scale has already been reported (Maccaferri et al., 2005; Somers et al., 2007). In this study, association mapping was applied to a set of durum elite accessions established within IDuWUE, an EU-funded project; the results concerning the genetic

structure of this set of accessions, the extent of LD as assessed with SSRs and the results of the association between agronomic and morpho-physiological traits and SSR markers are reported.

## MATERIALS AND METHODS

### *Plant material and field trials*

The collection of 189 durum wheat accessions includes mainly cultivars (cvs.) and advanced breeding lines from Mediterranean countries, CIMMYT and ICARDA. Most of the accessions are semi-dwarf elite materials released in the past four decades; a limited number of ‘founder accessions’ are also included. Fifteen trials exploring a broad range of water regimes were carried out in Italy, Spain, Morocco, Tunisia, Syria and Lebanon using an unreplicated modified augmented design with three checks. The following traits were considered: heading date (HD), plant height (PH), peduncle length (Pd), grain yield (GY), 1,000 grain weight (TKW), test weight (TWT) and mean number of grains per square meter (GRM2, calculated from GY and TKW).

### *Molecular characterization*

The durum collection was profiled with 186 genomic SSRs. The map position of SSR loci used for AM was obtained using an integrated map based on two durum RIL populations (Kofa × Svevo, Maccaferri et al., 2008; Colosseo × Lloyd, Mantovani et al. 2008, submitted).

The SSR data set was used to estimate LD values and significance ( $P$ ) among loci based on 10,000 permutations for all pairs of SSR loci. Rare alleles, i.e. with a frequency  $< 0.05$ , were excluded from the analysis.  $D'$  and  $r^2$  LD measures for loci with multiple alleles were used (Farnir et al., 2000). A subset of 96 SSRs loosely linked, evenly spread on the genome and showing null or limited LD values as to each other was selected to investigate the germplasm collection genetic structure using the Bayesian model-based clustering method implemented in the software STRUCTURE v. 2.1 (<http://pritch.bsd.uchicago.edu/structure.html>). The relatedness among accessions was further estimated using TASSEL and NTSYS programs. Association mapping was carried out using both the General Linear Model (GLM) and the Mixed Linear Model (MLM) in TASSEL.

## RESULTS AND DISCUSSION

### *LD and population structure*

The genetic relationships among the accessions were investigated using both a genetic-similarity and a model-based Bayesian clustering method. Both methods pointed out a high number of hypothetical subgroups in the elite durum panel, with a minimum number of five well distinct subgroups based on the origin/pedigrees of the accessions. In the Bayesian clustering, an increase of the number of subgroups led to improvements in the overall posterior probability; however, a generalized decrease of the number of accessions assigned to a specific subgroup with high posterior probability (e.g. > 0.50 or > 0.80) was observed. The five main subgroups were also clearly identified with the distance-based dendrogram and corresponded to: *i*) ICARDA (Omrahi) and Italian accessions originated from *syriacum* durum types (Haurani and Eiti), adapted to the most dryland areas, *ii*) advanced materials bred at ICARDA and in Italy for temperate areas and characterized by high yield potential, with Cham1 and Creso being the main representative accessions, *iii*) Italian cvs. related to the founders Valnova and Mexicali 75 (Stork), *iv*) cvs. and advanced lines bred at ICARDA, Italy, Spain and other Mediterranean countries from the CIMMYT cross Jori/Anhinga/Flamingo, *v*) accessions from CIMMYT, ICARDA, Italy, Spain and Morocco, all related to the CIMMYT founder Gallareta (= Altar 84), characterized by high yield potential. Average proportion of memberships of accessions to each of the five groups was equal to 0.11, 0.31, 0.29, 0.16 and 0.13, respectively.

Linkage disequilibrium (LD) pattern was investigated based on 186 SSR loci. The non syntenic pairs of loci with significant LD were 27.8% at  $P \leq 0.01$  (pair-wise error) and 15.1% at  $P \leq 0.001$ ; this disequilibrium significance was thus largely influenced by the presence of population structure in the germplasm collection. The LD  $r^2$  and  $D'$  estimates for the non syntenic loci were equal to 0.02 and 0.23, respectively. Pairs of physically linked (syntenic) loci showed LD values and significance largely influenced by the genetic linkage. In fact, most of the 140 pairs of loci with inter-marker distance comprised within 10 cM showed disequilibrium (83.6% of these marker pairs showed significant LD at  $P \leq 0.01$ , with an average  $r^2$  and  $D'$  values equal to 0.15 and 0.54, respectively). Wide LD presence was also observed for the 149 marker pairs with loci located between 10 to 20 cM. Due to the presence of significant LD effects caused by population structure, 'ad hoc' critical LD thresholds for the LD estimates between syntenic loci were thus obtained from the distribution of the non syntenic loci, according to Brescighello and Sorrells (2006). The 95<sup>th</sup> percentiles of these distributions, with  $r^2$  and  $D'$  values equal to 0.06 and 0.31, respectively, were thus used as critical thresholds to better discriminate the LD most likely caused by genetic linkage. The average extent of significant LD was estimated between 5 (based on  $r^2$  values) and 10 cM (based on  $D'$  values).

### *Association mapping*

The accessions were tested in 15 environments (envs.) highly differentiated as to thermo-pluviometric conditions with the final mean GY of accessions ranging from a minimum of ca. 10 q/ha to a maximum of 68 q/ha, with five envs. with low to medium production (10 to 40 q/ha), seven envs. with medium to high production (40 to 58 q/ha) and three envs. with GY higher than 60 q/ha. Correlations among envs. were frequently highly significant for HD, PH, Pd and TKW, with absolute  $r$  values comprised between 0.4 and 0.7; these traits showed a level of interaction with the environmental conditions lower than showed by GY (yield data from the different envs. were seldom correlated to each others) and, in part, yield components related to plant fertility.

As a result of these observations, AM results (data not reported) showed that, while for HD, PH, Pd and TKW it was possible to find a number of marker-trait associations that were repeatable across envs. (from four up to 11 envs.), for GY and, to a lesser extent, for GRM2 and TWT, the majority of marker-trait associations reached the significance in only two to four envs.

In general, when two or more markers were genotyped within a genetic distance equal or less than 10 cM, the marker-trait association results were in agreement. In some cases, AM results could be interpreted on a physiological basis by comparing the direction of the allelic effects across the traits that resulted significantly associated to a marker/chr. region. Two major effects on yield components (particularly TKW) and, to a lesser extent, HD and Pd, were observed at two regions of chr. 5A. The first region (near to centromere) was tagged by markers *Xwmc489.1*, *Xbarc303* and *Xgwm293*, while the second relevant region was identified by marker *Xgwm1570* (5AL chr.). Table 1 reports the mean values of the two allelic classes for the traits showing significant marker-trait association at the *Xgwm1570* locus. It should be noticed that the above mentioned associations were confirmed by mapping results obtained from field evaluations of two independent durum RIL populations (Colosseo  $\times$  Lloyd and Meridiano  $\times$  Claudio, see this conference Maccaferri et al., 2008; Mantovani et al, 2008); furthermore, these results were supported by the results of Brescighello and Sorrells (2006). Other chr. regions with significant effects on combinations of traits such as HD (and/or PH and Pd), yield components and, in some cases, GY, were found on chr. arm 1BL (*Xcfd65*, significant for PH, Pd, GRM2 and TKW; *Xcfa2086*, significant for HD, GRM2 and GY), 2AL (*Xgwm294*, significant for HD, PH, Pd, GRM2 and TKW), 2B centromeric (*Xgwm410*, significant for TKW and GY), 4BS (*Xbarc193*, *Xgwm1278* and *Xgwm856*, significant for HD, PH, GRM2, TKW and GY), 6AL (*Xwmc553*, significant for HD, PH, Pd, TKW and GY), 6B centromeric (*Xgwm518* and *Xbarc14*, significant for HD, GRM2, TKW, TW and GY), chr. 7AL and 7BL, distal homoeologous regions (*Xcfa2123*, *Xwmc273* and *Xcfa2257* on chr. 7AL and *Xwmc276*, *Xcfa2040*, *Xwmc526* and *Xcfa2257* on chr. 7BL), associated to PH, Pd, GRM2, GY and TWT.

In general, the number of frequent alleles/SSR locus used in the association test ranged from two to four. As to loci with three to four alleles, box plots (data not shown) and least significant differences showed that, in most cases, the effects were due to one allele only.

The phenotypic and molecular characterization of the germplasm assembled within the IDuWUE project proved to be useful allowing for a better investigation of the genetic bases of important agronomic, morpho-physiologic and disease resistance traits in durum wheat.

Table 1. Mean values of the two allelic classes at the *Xgwm1570* locus for 1,000 grain weight (TKW), test weight (TWT) and number of grains per square meter (GRM2) in the 15 Mediterranean environments.

<i>Xgwm1570</i>	Environments <sup>a</sup>														
	Tns	Itl-Cd	Spn-Gs	Itl-Cr	Syr	Itl-Cr	Lbn	Syr	Tns	Lbn	Lbn	Mrc	Mrc	Lbn	Spn-Gr
Traits	i05	r04	r04	r04	i05	r05	i05	r05	r05	i04	r05	i04	r04	r04	r05
Mean yield (q/ha)	67.8	64.8	64.5	57.9	58.6	56.6	56.1	46.4	43.8	40.9	36.0	35.9	30.0	22.1	10.0
TKW (g/1000 seeds)															
Allele 1 (77 counts)	33.3	<b>44.1</b>	<b>45.5</b>	40.8	38.3	<b>32.8</b>	<u>45.8</u>	36.6	27.4	<b>37.1</b>	<b>35.2</b>	<u>39.0</u>	<u>38.9</u>	37.8	27.6
Allele 2 (90 counts)	33.1	<b>41.8</b>	<b>43.5</b>	<u>39.2</u>	37.3	<b>31.1</b>	<u>44.5</u>	35.7	26.8	<b>35.9</b>	<b>34.4</b>	<u>38.7</u>	<u>37.8</u>	37.6	27.4
GRM2 (no./m <sup>2</sup> )															
Allele 1	20537	<b>15647</b>	<b>14300</b>	<u>14150</u>	15258	<u>17613</u>	<u>12307</u>	12292	16641	<u>11406</u>	10382	8744	7190	5950	3339
Allele 2	20929	<b>16247</b>	<b>14913</b>	<u>14510</u>	15582	<u>18142</u>	<u>12620</u>	12522	16749	<u>11685</u>	10489	8837	7399	6077	3353
Pd (cm)															
Allele 1	<b>18.2</b>	17.8	20.6	18.8	19.8	13.9	<u>15.6</u>	14.8	17.6	14.4	<u>11.8</u>	-	-	9.4	5.6
Allele 2	<b>19.6</b>	17.9	21.2	18.9	20.5	14.2	<u>16.4</u>	15.4	17.8	14.4	<u>12.3</u>	-	-	9.3	5.6

a) Environments: Tns: Tunisia, Kef; Itl-Cd: Italy, Cadriano; Spn-Gs: Spain, Gimenells; Itl-Cr: Italy, Cerignola; Syr: Syria, Aleppo; Lbn: Lebanon, Rayack; Mrc: Morocco, Sidi Elaydi; Spn-Gr: Spain, Granada.

Underscored values indicate a significant ( $P \leq 0.05$ ) difference between the two allelic classes.

Bold underscored values indicate a significant ( $P \leq 0.01$ ) difference between the two allelic classes.

**Acknowledgements.** The financial contribution of the European Union (IDuWUE project; INCO contract number ICA3-CT-2002-10 028) is gratefully acknowledged. The authors thank dr. Sandra Stefanelli for the technical assistance in field trials throughout the project.

## REFERENCES

- Breseghello F, Sorrells ME (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172: 1165-1177.
- Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. *Curr Opin Plant Biol* 5: 107-111.
- 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.
- Maccaferri M, Sanguineti MC, Corneti S, Araus Ortega JL, Ben Salern M, Bort J, DeAmbrogio E, del Moral LG, Demontis A, El-Ahmed A, Maalouf F, Machlab H, Martos V, Moragues M, Motawaj J, Nachit MM, Nserallah N, Ouabbou H, Royo C, Slama A, Tuberosa R (2008). Quantitative trait loci for grain yield and adaptation of durum wheat (*Triticum durum* Desf.) across a wide range of water availability. *Genetics* 178: 489-511.
- Maccaferri M, Sanguineti MC, Noli E, Tuberosa R (2005). Population structure and long-range linkage disequilibrium in a durum wheat elite collection. *Molecular Breeding* 15:271-290.
- Mantovani P, Maccaferri M, Sanguineti MC, Tuberosa R, Catizone I, Wenzl P, Thomson B, Carling J, Eric Huttner E,
- DeAmbrogio E, Kilian A (2008). An integrated DArT-SSR linkage map. *Mol Breed*, in press.
- Morgante M, Salamini F (2003). From plant genomics to breeding practice. *Curr Opin Biotech* 14: 214-219.
- Somers GJ, Banks T, Depauw R, Fox S, Clarke J, Pozniak C, McCartney C (2007). Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. *Genome* 50: 557-567.
- Varshney RK, Tuberosa R (2007). Genomics-Assisted Crop Improvement: an overview. In RK Varshney, R Tuberosa, eds, *Genomics-Assisted Crop Improvement*, Vol. 1: *Genomics Approaches and Platforms*, Springer, Dordrecht, pp 1-12.