Theor Appl Genet (2014) 127:791–807
DOI 10.1007/s00122-013-2257-8

ORIGINAL PAPER

Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes

Erena A. Edae · Patrick F. Byrne · Scott D. Haley · Marta S. Lopes · Matthew P. Reynolds

Received: 29 July 2013 / Accepted: 14 December 2013 / Published online: 10 January 2014

Abstract
Key message **A stable QTL that may be used in marker-assisted selection in wheat breeding programs was detected for yield, yield components and drought tolerance-related traits in spring wheat association mapping panel.**
Abstract Genome-wide association mapping has become a widespread method of quantitative trait locus (QTL) identification for many crop plants including wheat (*Triticum aestivum* L.). Its benefit over traditional bi-parental mapping approaches depends on the extent of linkage disequilibrium in the mapping population. The objectives of this study were to determine linkage disequilibrium decay rate and population structure in a spring wheat association mapping panel (n = 285–294) and to identify markers associated with yield and yield components, morphological, phenological, and drought tolerance-related traits. The study was conducted under fully irrigated and rain-fed conditions at Greeley, CO, USA and Melkassa, Ethiopia in 2010 and 2011 (five total environments). Genotypic data were generated using diversity array technology markers. Linkage disequilibrium decay rate extended over a longer genetic distance for the D genome (6.8 cM) than for the A and B genomes (1.7 and 2.0 cM, respectively). Seven subpopulations were identified with population structure analysis. A stable QTL was detected for grain yield on chromosome 2DS both under irrigated and rain-fed conditions. A multi-trait region significant for yield and yield components was found on chromosome 5B. Grain yield QTL on chromosome 1BS co-localized with harvest index QTL. Vegetation indices shared QTL with harvest index on chromosome 1AL and 5A. After validation in relevant genetic backgrounds and environments, QTL detected in this study for yield, yield components and drought tolerance-related traits may be used in marker-assisted selection in wheat breeding programs.

Communicated by A. Grane.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-013-2257-8) contains supplementary material, which is available to authorized users.

E. A. Edae (✉) · P. F. Byrne · S. D. Haley
Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523, USA
e-mail: erenakae@yahoo.com

M. S. Lopes · M. P. Reynolds
CIMMYT, Int., Apdo. Postal 6-641, 06600 Mexico, DF, Mexico

Introduction

Wheat is the world's third most important food crop next to maize (*Zea mays* L.) and rice (*Oryza sativa* L.) (Green et al. 2012). It accounts for 19 % of total production among major cereal crops and provides 55 % of the carbohydrate consumed by humans around the world (Gupta et al. 1999; Bagge et al. 2007). However, its productivity is often reduced by both biotic and abiotic stresses and its potential yield is rarely achieved.

Plant breeding has successfully improved crop resistance to both biotic and abiotic stresses, including drought, through phenotypic selection (Araus et al. 2008; Cooper et al. 2009). However, the progress is slow, and there is a large yield gap between drought prone areas and ideal production regions for most crops, including wheat. Many previous studies have shown that tolerance to drought is a complex quantitative trait that involves multiple chromosome regions (Barnabas et al. 2008; Fleury et al. 2010; Pinto et al. 2010; Ravi et al. 2011; Mir et al. 2012). It is

further complicated by the fact that the degree of drought effects on plants depends on timing, duration and intensity of drought, and different traits may be required for different patterns of drought (Passioura 2012).

Genetic improvement under drought requires identifying sources of traits associated with drought tolerance and introgressing the genes underlying the target traits to locally adapted cultivars. The challenge for implementing this strategy in breeding programs is the identification of the most suitable target traits in a time-efficient and cost-effective way for different drought scenarios (Passioura 2012). Recent advancements in high throughput genotyping and phenotyping have improved understanding of the physiological and molecular bases underlying complex traits including drought tolerance (Collins et al. 2008; Habash et al. 2009; Mir et al. 2012; Sinclair 2012). QTL mapping is a key approach for understanding the genetic architecture of complex traits in plants (Holland 2007). However, QTL mapping using bi-parental populations explains only a small portion of the genetic architecture of a trait because only two alleles per locus can be evaluated at a time. Other limitations of bi-parental populations are low mapping resolution, population specificity of detected QTL, and the long time required to develop mapping populations. These limitations have partly contributed to the slow transfer of knowledge from bi-parental QTL studies to practical applications in plant breeding.

The advent of the association mapping approach has overcome some of the limitations of bi-parental mapping populations. Since association mapping utilizes diverse germplasm, QTL for many traits can be detected with high resolution in a single study, making the method more efficient and less expensive than bi-parental QTL mapping (Breseghello and Sorrells 2006; Ersoz et al. 2009; Sorrells and Yu 2009; Waugh et al. 2009). The resolution and power of association studies, however, depend on the extent of linkage disequilibrium (LD) across the genome. LD needs to be determined in each study as it is affected by several factors such as population history, recombination frequency and mating system.

The correlation of allele frequency (r^2) among the markers is the common statistic used to measure LD (Gupta et al. 2005; Sorrells and Yu 2009). LD is expected to decay as a function of the nucleotide or linkage distance, as recombination reduces LD. This guides decisions on the number of markers required to conduct association mapping in a crop species (Waugh et al. 2009). To visualize LD patterns and the rate of LD decay for a chromosome, r^2 values are usually plotted against nucleotide or linkage distance (Abdurakhmonov and Abdukarimov 2008).

Previous studies have demonstrated unique LD patterns for different crop species and populations within a species, with rapid levels of LD decay observed in cross-pollinated species (e.g., maize) compared to self-pollinated species (e.g., wheat) (Wilson et al. 2004; Chao et al. 2007; Comadran et al. 2009). Although association mapping has advantages over bi-parental populations, QTL identification could be confounded by population subgroups and plant phenology. Another limitation is that markers with low allele frequencies are often not considered in association analysis. However, statistical models have been developed to account for population structure and familial relationship among the genotypes in the mapping panel (Yu et al. 2006).

Association mapping has been used successfully to detect QTL in wheat for disease resistance (Crossa et al. 2007; Maccaferri et al. 2010; Yu et al. 2011, 2012; Adhikari et al. 2012; Kollers et al. 2013a, b), end-use quality traits (Breseghello and Sorrells 2006; Zheng et al. 2009), Russian wheat aphid (*Diuraphis noxia*) resistance (Peng et al. 2009), and yield and yield component traits (Maccaferri et al. 2011; Neumann et al. 2011). The suitability of diversity array technology (DArT) markers for association studies has been proved particularly for species lacking cost-effective single-nucleotide polymorphism (SNP) markers (Benson et al. 2012). However, the majority of previous studies have been conducted either with low marker density or a small number of lines in the mapping population. Therefore, the objectives of the present study were to (1) determine LD decay rate in a spring wheat association mapping panel, (2) analyze population structure in the panel, and (3) identify markers associated with yield and yield components, morphological, phenological and drought tolerance-related traits.

Materials and methods

Mapping population

The spring wheat association mapping panel used in this study wheat association mapping II (WAMII) was originally developed by the International Maize and Wheat Improvement Center (CIMMYT) with the intention of identifying QTL/genes for drought and heat tolerance. The panel comprised a total of 287 diverse lines which were assembled from the Elite Spring Wheat Yield Trial (26th, 27th and 28th ESWYT), Semiarid Wheat Yield Trial (1st–16th SAWYT) and High Temperature Wheat Yield Trial (HTWYT) (Lopes et al. 2012). Many synthetic hexaploid-derived wheat lines were included in the panel (Lopes and Reynolds 2012). In the study reported here, from 283 to 294 lines were evaluated depending on the location and year due to limitations in seed quantity.

Experimental design and phenotypic trait evaluation

In 2010, association mapping panel (including two local check cultivars, Reeder and Butte 86 (Mergoum et al.

2006)) was evaluated under fully irrigated conditions at the USDA-Agricultural Research Service Limited Irrigation Research Farm in Greeley, CO, USA (latitude 40° 27′N; longitude 104° 38′W; elevation 1,427 m). The soil at the site is well drained with fine sandy loam to clay loam texture and a pH range of 7.4–8.4. The meteorological data with amount of water supplemented by irrigation are summarized in Table 1 for all environments.

In 2011, we evaluated the mapping panel [including two local checks, SD3870, a breeding line form South Dakota and Granger (Glover et al. 2006)] at Greeley under both fully irrigated and rain-fed conditions. The irrigated treatment was supplemented three times with drip irrigation, (twice before flowering and once during the grain filling stage), while the rain-fed treatment was irrigated only once at flowering to avoid complete failure of the experiment.

In the 2010 and 2011 experiments at Greeley, the lines were planted in four-row plots 1.52 m wide and 0.92 m long with 0.20 m spacing between rows and a seeding rate of approximately 173 seeds m^{-2}. Each entry was replicated twice in a Latinized incomplete row–column design with CycDesign 3.0 software (www.cycdesign.co.nz). The experimental field was maintained free of weeds by manually removing weeds as required. In both seasons, the plants experienced heat stress mainly from heading through physiological maturity, as maximum temperatures were over 30 °C for a total of 13 days in June and 22 days in July 2010; temperatures exceeded 30 °C for 15 days in June and 27 days in July 2011.

The association mapping panel was planted at the Melkassa Agricultural Research Center of the Ethiopian Institute of Agricultural Research, Melkassa, Ethiopia (latitude 8° 24′N; longitude 39° 21′E, elevation 1,550 m), on wet soil from rain in the previous few days for non-stressed treatment. The same set of lines was planted on drier soil for stressed treatment. The dominant soil type at Melkassa is sandy loam (Andosol of volcanic origin) with pH ranging from 7.0 to 8.2. The experiment was laid out as an alpha-lattice design with 14 entries per incomplete block and two replications. A two-row plot of length 2.5 m, width 0.4 m and between row spacing of 0.20 m was used. Seeding rate was based on the local recommendation of 150 kg ha^{-1}. Nitrogen fertilizer was applied in split doses at planting and tillering at a rate of 50 kg ha^{-1} for each dose. Phosphorus fertilizer was applied as diammonium phosphate at planting at the rate of 100 kg ha^{-1}. The temperature was below 30 °C for all days from emergence time through physiological maturity.

The phenotypic traits evaluated in this study are defined as follows. Plant height was recorded as the average of three values for each plot measured in centimeter from the soil surface to the tip of the spike excluding awns. Days to heading were recorded as the number of days from planting until 50 % of the spikes in each plot had completely emerged above the flag leaves. Days to maturity were recorded as the number of days from planting until 50 % of the peduncles in each plot had turned yellow. Grain filling duration was calculated as the difference between the days to heading and days to maturity. Normalized vegetation index (NDVI) was obtained by scanning plants in each plot at the grain filling stage with a GreenSeeker instrument model 3541 (NTech Industries Inc., Boulder, CO). Green leaf area was obtained from a photo taken at a height of approximately 0.5 m directly above each plot with a digital camera (Coolpix S8100, Nikon Corp., Japan) during vegetative stage, and pictures were processed with Breedpix software (Casadesus et al. 2007). Leaf senescence was scored on a scale from 0 to 10, where 0 indicates completely green leaves and 10 indicates that all leaves in a plot had changed completely to yellow. Flag leaf length (measured from the leaf collar to the tip) and width (measured at the widest part of the flag leaf) were recorded as the average measurement of three flag leaves per plot. Flag leaf area (cm^2) was calculated as flag leaf length × flag leaf width × 0.75.

Single-kernel diameter (mm), kernel hardness and single-kernel weight (mg) were determined from 100 seeds (sampled from grain of the biomass sample) in a single-kernel characterization system instrument Model 4100 (Perten Instruments, Springfield, IL). Spike length, spikelet number per spike, kernel number and weight (g) per spike, and kernel number per spikelet were recorded as the average of five spikes per plot. Thousand-kernel weight

Table 1 Meteorological data including total rainfall, average maximum and average minimum temperature (avg max/avg min), and total amount of water supplied by irrigation during growing seasons of spring wheat association mapping panel (WAMII) in 2010 and 2011

Country	Environment	Lines evaluated	Planting date	Total rainfall (mm)	Irrigation (mm)	Avg max/avg min (°C)
USA	GRW10	285	April 5	271	93.8	16.94/0.22
	GRW11	288	April 15	173	140	17.6/0.83
	GRD11	288	April 15	173	19	17.6/0.83
Ethiopia	MLKW11	294	July 17	533	0	27.3/8.5
	MLKD11	294	July 19	533	0	27.3/8.5

was determined by extrapolation after counting seeds of five spikes with a seed counter (International Marketing and Design Corp Model 900-2; San Antonio, TX) and obtaining the weight of the seeds. Number of spikes m^{-2} was calculated by dividing the number of kernels m^{-2} by kernel number per spike. The number of kernels m^{-2} was obtained from the ratio of grain weight m^{-2} to thousand-kernel weight, multiplied by 1,000. Biomass weights were obtained by cutting all the plants at ground level in one row of each plot at maturity, then weighing samples after 48 h in a 40 °C drier. These samples were threshed to obtain grain weight, and harvest index was recorded as the ratio of grain weight to oven-dried biomass. Grain yield was the total weight of seed yield in each plot divided by the plot area and expressed as kg ha^{-1}.

Test weight (kg hL^{-1}) was determined using standard procedures from a small sample of the grain collected at harvest. Drought susceptibility index (DSI) was calculated using grain yield and kernel number per square meter under irrigated and rain-fed conditions as described by Fisher and Maurer (1978). DSI = $(1-Y_d/Y_i)$/DII, where Y_d = yield of each line in the dry treatment, Y_i = yield of each line under fully irrigated conditions and DII = $1-(Y_{dm}/Y_{im})$ where Y_{dm} is the average yield of the dry treatment and Y_{im} is the average yield of the irrigated treatment.

Phenotypic data analysis

The phenotypic data analyses were conducted with SAS v. 9.3 software (SAS Institute Inc., Cary, NC). First, the general linear model (GLM) procedure was used to obtain best linear unbiased estimates, considering genotype, replications, rows and columns as fixed in the model for each environment. Normality of the data for each trait was checked using a Q–Q plot of residuals in the SAS GLIMMIX procedure. The presence of statistically significant differences among the genotypes for each trait was also checked with the GLM procedure. Then, best linear unbiased predictions (BLUPs) and variance components were obtained for all traits using a Mixed model procedure, considering all factors in the model as random. Environment was considered fixed in the combined data analysis. To account for spatial variations in the experimental field, four spatial variability adjustment models (spatial power, anisotropic spatial power, Matérn spatial and autoregressive) were tested for each trait. The correlation values due to spatial variability in each model were found to be very low for all data sets except for Greeley in 2010. Thus, the autoregressive spatial adjustment model was applied for the data set in 2010, but no adjustment was made for the remaining environments.

Genotypic data analysis

Diversity array technology marker genotypes were obtained following the procedures of Akbari et al. (2006) at Triticarte Pty. Ltd. (Canberra, Australia; http://www.triticarte.com.au), a whole-genome profiling service laboratory. A total of 1,863 DArT markers were used in the analyses, after markers with <5 % allele frequency and those with a high percentage of missing data points (>6 %) were removed. Genome-wise distribution of the markers was 558 on genome A, 617 on genome B, and 290 on genome D (http://www.triticarte.com.au). Chromosome map positions were not known for 398 markers. A DArT marker physical map (based on Chinese spring wheat deletion lines) (http://www.cerealdb.uk.net/) was used to assign trait-associated markers to chromosome arms.

Population structure and linkage disequilibrium analyses

Seventy-eight markers (3–4 markers spaced >10 cM per chromosome) were selected from all chromosomes (except for chromosome 4D and 5D) from a total of 1,863 markers for analysis of population structure. To determine population structure, an admixture model with correlated allele frequency in STRUCTURE software was applied (Pritchard et al. 2000). A burn-in of 20, 000 iterations followed by 20,000 Monte Carlo Markov Chain (MCMC) replicates was conducted to test k values (number of subpopulations) in the range of 3–12. Each k was replicated five times and the run that assigned the most lines with probability >0.5 in all clusters was used. The likely number of subpopulations was determined using the approach of Evanno et al. (2005) and the likelihood distribution of k was examined. Genetic distance-based cluster analysis was conducted using hclust script for Ward method in the R package (www.cran.r-project.org) using the same 78 markers to compare results with STRUCTURE software output. Multiple regression analysis was also done for all phenotypic traits using population subgroups in the model to determine the extent of the confounding effect of population structure on the phenotypic traits.

LD among markers was calculated using observed versus expected allele frequencies of the markers in TASSEL v.3.0 (Bradbury et al. 2007). Only mapped markers were used for LD calculation both for the panel and for model-based subgroups. The critical r^2 value beyond which LD is due to true physical linkage was determined by taking the 95th percentile of the square root transformed r^2 data of unlinked markers (Breseghello and Sorrells 2006). The percentage of marker pairs significant at different critical r^2 values (0.2 and 0.2641) and $P < 0.001$ was determined for each chromosome to compare the degree of LD among

chromosomes. Locally weighted polynomial regression (LOESS)-based curves were fitted on scatter plots of r^2 versus distance among markers. LOESS is a non-parametric method of estimating local regression surfaces, and it is a robust fitting method particularly when there are outliers in the data (Cleveland 1979). The LOESS model is written as $y_i = g(x_i) + \varepsilon_i$, where y_i is ith measurement for a response variable y, x_i is the corresponding measurement of a predictor variable x, ε_i is a random error and g is the regression function. Analysis of molecular variance was conducted using the seven groups with Arlequin software (http://cmpg.unibe.ch/software/arlequin3).

Marker–trait association (MTA) analysis

A total of 1,863 high-quality DArT markers were used in this study. The GAPIT program in R software (Lipka et al. 2012) was applied to determine the association between markers and phenotypic traits. MTA analysis was conducted for each environment separately, using BLUPs for each trait and line. A mixed linear model was employed by including BLUPs, marker, kinship matrix (K) and probability of membership of each line (Q) in the model for each trait (Yu et al. 2006). Kinship matrix and principal component analyses were conducted in TASSEL software.

Model comparison was made for K (kinship) only, Q + K (population structure and kinship) and P + K (principal component and kinship) models. Mean square of the difference (MSD), based on observed P values and expected P values, and Q–Q plot were used to compare the models; MTA P values for yield of five environments plus the combined data set were used for the model comparisons. Among the three mixed models, the model taking into account population structure and genotype relationship showed the least deviation from the nominal alpha level in most cases. The same procedures were followed to compare mixed model versus GLM model, and the former was found better in controlling false positives. For multiple comparison adjustment, false discovery rate (FDR)-adjusted P values were calculated for each trait (Benjamini and Hochberg 1995). However, many markers within 10 cM distance are in LD, and FDR adjustment is still too stringent as it assumes independent testing. The proportion of cross-validated phenotypic variations explained by marker–trait associations that survived FDR adjustment were calculated using cvGWAS R package through fivefold cross-validation procedures as outlined by Shen (2013). The mapping panel was split into a training data set comprising 80 % of the samples and a validation data set consisted 20 % of the panel to fit a linear regression of the phenotype on the genotype for each marker in the training set, and then predict phenotypic values in the validation sets using an estimated model from the training set. The mean of the squared correlation coefficients (R^2) between observed phenotypic measurements and their predicted values in each of the five validation sets was taken as an estimate of the proportion of phenotypic variance explained by the marker.

Results

Wide differences among the study materials were visually observed for many traits during field evaluations. Water deficit reduced full expression of those traits in rain-fed treatments despite increase of expression for some drought-related traits (e.g., leaf waxiness). The rain-fed treatments at Melkassa experienced water deficit only at emergence stage, whereas at Greeley the rain-fed treatment was exposed to water deficit starting from vegetative stage through grain filling. Data were collected for a total of 26 traits, but this number varied depending on the year and location.

Agronomic trait means

Analysis of variance showed significant differences ($P < 0.05$) among genotypes for most traits in all environments and for the combined data analysis across environments. The mean grain yield of individual lines in the five environments were ranged from 1,087 kg/ha (recorded at Greeley in 2011 under rain-fed conditions) to 5,377 kg/ha (obtained at Melkassa under non-stressed conditions). The mean grain yield (2,156 kg/ha) recorded under fully irrigated conditions in 2010 was the highest of the three trials grown in Greeley (Online Resource 1). The vegetative stage of the lines (calculated as the number of days from planting to heading) was longer in the Greeley environments (mean of 68 days) than in the Melkassa environments (mean of 55 days). However, the grain filling duration at Melkassa was longer than that of the Greeley environments (39 vs. 34 days). On average, the genotypes headed 13 days earlier at Melkassa than at Greeley. Heading date ranged from 47 to 69 days at Melkassa and from 63 to 72 days at Greeley (Online Resource 1; Online Resource 2). In Melkassa environments plants grew taller than in the Greeley environments.

Model-based population structure and linkage disequilibrium

Population structure analysis of 287 spring wheat lines conducted with the STRUCTURE program indicated that the likely number of subpopulations was seven based on change of k. Of these, subpopulations II, IV, V and VI were dominated by the lines with Kauz, Pastor, TUI and

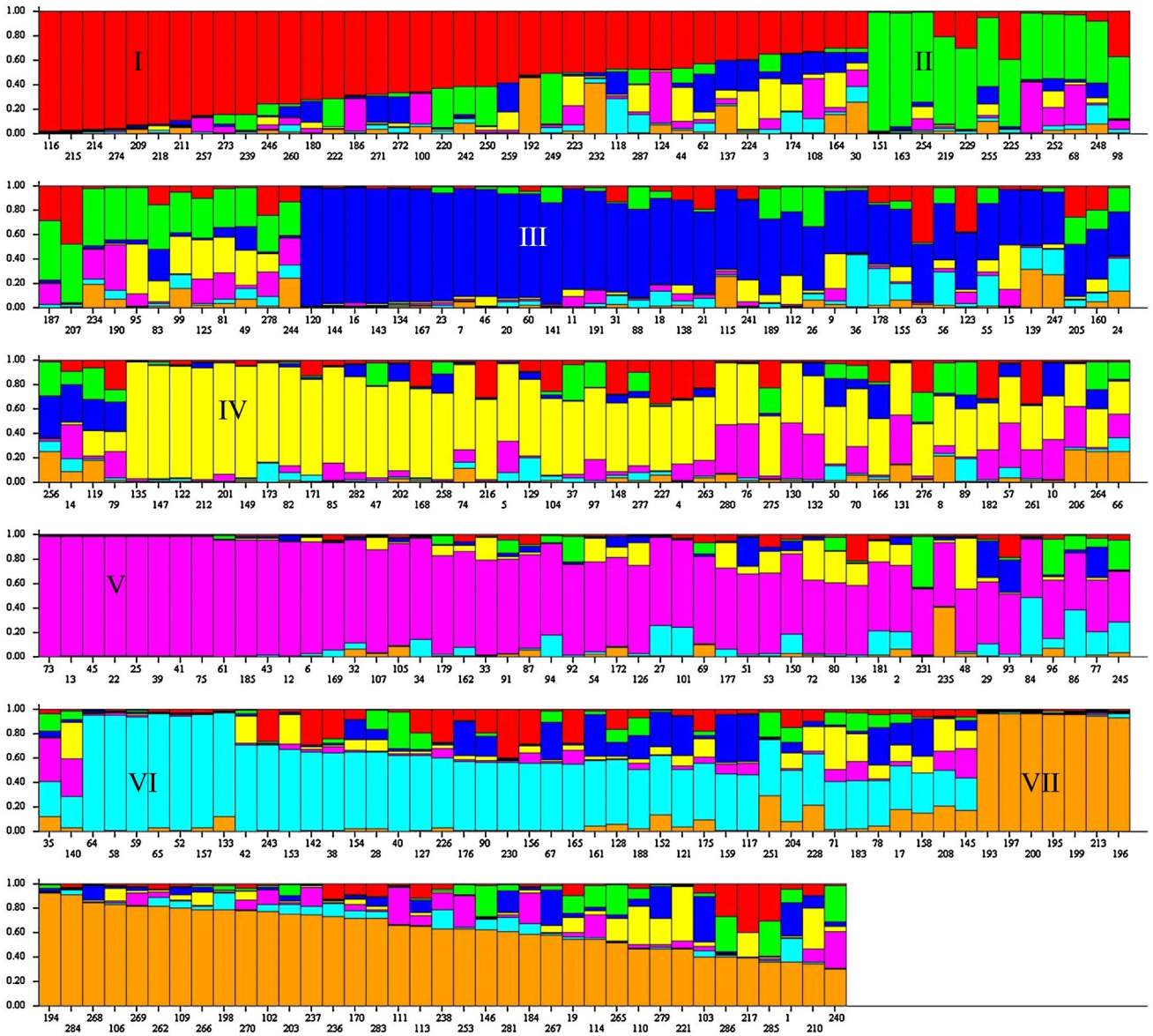

Fig. 1 Population structure for 287 entries in a spring wheat association mapping panel based on 78 DArT markers

WBLL1 background, respectively (Fig. 1). Lines with different backgrounds were grouped together for subpopulations I, III and VII. Genetic distance-based cluster analysis also provided evidence for the presence of subpopulations despite the lack of similarity between its clusters and the subgroups of model-based analysis in STRUCTURE. Molecular variance analysis for the seven model-based populations indicated that 78.5 % of the total variation is explained by within-population variation, whereas 21.5 % of the variation is due to among-population variation (data not shown). Population differentiation (Fst) values ranged from 0.14 to 0.73 and were highly significant ($P < 0.0001$) for all pairs, supporting the presence of population structure.

Linkage disequilibrium among markers was calculated for all chromosomes (except chromosomes 4D and 5D that were represented by only a single marker each). A critical value of $r^2 > 0.264$ was determined to be the appropriate threshold for LD due to physical linkage. Chromosomes 4A (62 %) and 1B (55 %) showed a higher percentage of significant ($P < 0.01$) marker pairs in LD whereas chromosomes 5A (20 %), 2B (23 %) and 7A (23 %) had the least number of significant ($P < 0.01$) marker pairs (Online Resource 3). The percentage of marker pairs due to physical linkage was high for chromosome 3D (24 %), followed by chromosomes 2D (17 %) and 1B (16 %). The percentage of LD due to physical linkage mimics the percentage of LD at $r^2 > 0.2$ for all chromosomes, but had no similarity

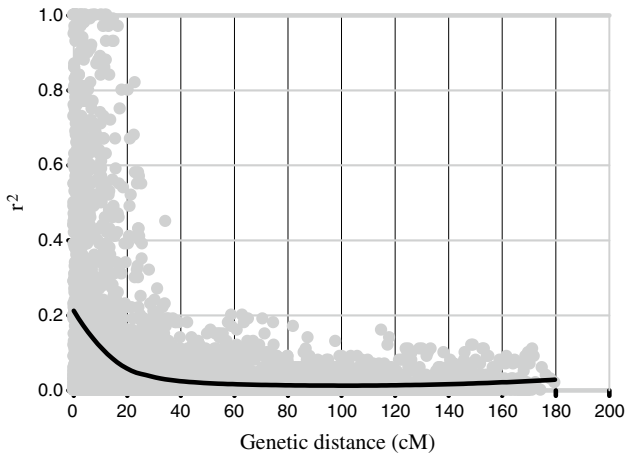

Fig. 2 Linkage disequilibrium (r^2) plot of all chromosomes of the A genome in 287 lines of a spring wheat association mapping panel

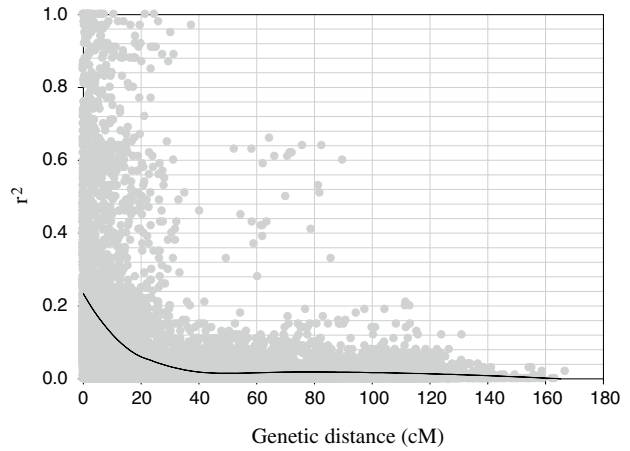

Fig. 3 Linkage disequilibrium (r^2) plot of all chromosomes of the B genome in 287 lines of a spring wheat association mapping panel

with that of LD at $P < 0.01$. Marker pairs at $r^2 > 0.2$ and $r^2 > 0.264$ were significant at $P < 0.001$ for all 19 chromosomes considered in this study.

Linkage disequilibrium decay rate evaluation was conducted at the genome and individual chromosome level. The genome level LD decayed below $r^2 = 0.2$ at about 1.7 cM for the A genome (Fig. 2), while the smoothing curve crossed the $r^2 = 0.2$ line at approximately 2 cM for the B genome (Fig. 3). For the D genome, the curve crossed the $r^2 = 0.2$ line near 6.8 cM genetic distance (Fig. 4). For all 19 chromosomes, the LD decay curve crossed the $r^2 = 0.2$ line at about 3.4 cM. We were able to determine the genetic distance at the baseline $r^2 = 0.2$ for four out of the seven model-based subgroups for all chromosomes together; LD decayed below $r^2 = 0.2$ within 8–9 cM for three of them and within 6 cM for the fourth subpopulation.

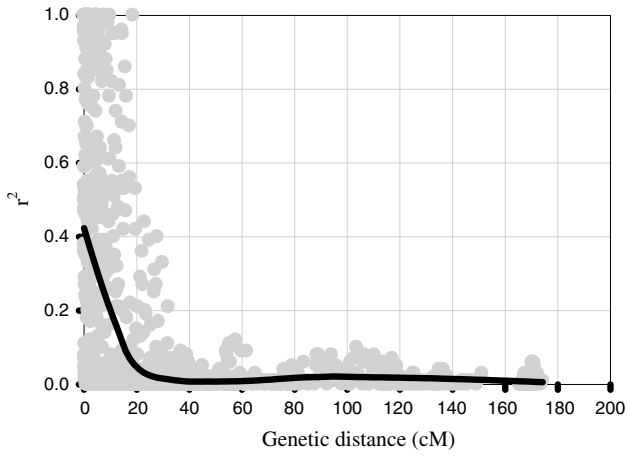

Fig. 4 Linkage disequilibrium (r^2) plot of all chromosomes on the D genome in 287 lines of spring wheat association mapping panel

Table 2 Percent of phenotypic variation explained (R^2) by population structure based on combined data across environments

Trait	Environments[a]	R^2 (%)	P value[b]
Grain yield	5	7.4	0.0015
Thousand-kernel weight	5	21.0	0.0001
Harvest index	5	2.2	0.3950
Kernel number	5	3.5	0.1182
Spikelet number	3	13.0	0.0001
Kernel weight per spike	3	5.3	0.020
Test weight	3	5.0	0.0212
Days to heading	5	4.7	0.0362
Days to maturity	5	3.0	0.1970
Grain filling duration	5	9.1	0.0001
Flag leaf area	5	13.2	0.0001
Kernel hardness	3	1.9	0.506
Single-kernel weight	3	25.9	0.0001
Kernel number per spikelet	3	9.0	0.0002
Spike number per square meter	3	5.6	0.014
Single-kernel diameter	3	20.5	0.0001
Kernel number per spike	3	15.7	0.0001
Flag leaf length	5	11.4	0.0001
Flag leaf width	5	9.4	0.0001
Plant height	4	29.5	0.0001
Biomass	5	12.4	0.0001
NDVI	3	18.9	0.0001
Green leaf area	5	14.4	0.0001

NDVI normalized difference vegetation index

[a] The number of environments used in combined data analysis

[b] Significance threshold P value is 0.05

Multiple regression analysis with population structure in the model showed that plant height (29.5 %) was the most affected by the genotype groupings, followed by

single-kernel weight (25.9 %), thousand-kernel weight (21.0 %), single-kernel diameter (20.5 %) and NDVI (18.9 %) (Table 2). On the other hand, the variations explained due to population structure were non-significant for kernel number per square meter, drought susceptibility index, kernel hardness, harvest index and days to maturity. Moreover, population structure explained only about 5 % of the total variation in days to heading, test weight and kernel weight per spike. The variation explained due to population structure in grain yield (7.5 %) was also low.

Marker–trait associations

Although several MTA were detected at $P < 0.05$ for all traits, we are reporting only strong MTA ($P < 0.001$) for single environments and moderate MTA significant ($P < 0.01$) in at least half of the test environments. Consistency across environments was used as an additional criterion for MTA significant at $P < 0.01$ to reduce the risk of including false marker–trait associations. A summary of MTA in different environments for each phenotypic trait is given in Table 3.

Considering both criteria together ($P < 0.001$ and $P < 0.01$ in half or more of the environments), a total of 565 MTA were detected in one or more environments for 26 measured or calculated phenotypic traits in five environments plus combined data across environments. Out of these, about 20 % of the MTA were detected only in a single environment and the remaining 80 % showed up in two or more environments. A total of 130 (22.9 %) of MTA involved unmapped markers. The numbers of MTA detected for grain yield both under irrigated treatment and rain-fed treatment at Greely were similar, while at Melkassa the number of MTAs detected for grain yield under the stressed treatment was lower than that detected under non-stressed treatment for grain yield (Table 3). The highest number of MTA was recorded for kernel hardness (113) followed by test weight (44) and flag leaf length (39) while the fewest MTA were obtained for drought susceptibility index, flag leaf senescence, kernel number per spikelet, kernel number per spike and spikelet number per spike. Moreover, kernel hardness had the largest number of stable MTA (15) followed by test weight (9). Chromosome-wise, the highest number of MTA was detected on chromosomes 5B, 3B, 7A and 1B, while chromosomes 1A, 2A, 2D, 3A and 5A harbored the fewest MTA in this study. No MTAs were detected for chromosomes 4D and 6D.

Grain yield MTAs were detected on chromosomes 1BS, 2DS, 5B (73 and 76.4 cM) and 7B. Unmapped marker *wpt0419* was also associated with grain yield at Melkassa under irrigated conditions. The marker *wpt6531* on chromosome 2DS was associated with yield in four out of six environments, including both irrigated and rain-fed conditions, and can be considered as a stable marker for grain yield. However, *wpt3457* (5B) showed the strongest association with yield under rain-fed conditions at Greeley in 2011 (Online Resource 4). Stable MTA were also detected for the major yield component traits kernel number per square meter on chromosome 7AS and harvest index on chromosomes 5AL, 5B (72.4 cM) and *wpt0286* (unmapped). Regions of chromosomes 1BL, 3BS, 4A, 5B (72.4 cM) and 5BL were consistently associated with thousand-kernel weight. Regions of chromosome 4B, 5B, 6B, 7AS and 7AL were associated with spike number per square meter at two environments. However, all MTA obtained for final biomass on chromosomes 1AS, 5B (72.4 cM), 7BL, 7D, 7DL and 7DS were environment specific (Online Resource 5). Similarly, all MTA detected for number of spikelets per spike (2B and 7B), kernel number per spikelet (1DL, 7A and 7BL) and kernel number per spike (1AS, 3BS and 7A) were detected only in single environments despite the presence of very strong associations for some MTA (Online Resource 4).

Single-kernel traits such as single-kernel weight and diameter and kernel hardness had more stable MTA than most of the yield component traits. The MTAs for single-kernel weight were distributed on chromosomes 1BL, 1D, 4A, 2AL, 4BL and 5BL while MTAs of single-kernel diameter were detected on chromosomes 1BL, 2D, 3AS, 3B, 3D, 4AL, 6BS, 7BL and 7DL. Several MTA were obtained for kernel hardness and the most stable ones (those detected in all environments) were found on chromosomes 1BL, 1D, 3AS, 3D, 4AL and 7A. Similarly, many stable MTA were obtained for test weight, with chromosomes 2DL, 3BS, 4A, 4BL and 7BL comprising the location of MTA detected in three out of the total four environments.

The most stable MTA for days to heading was detected on chromosome 1DS (four out of six environments), followed by MTA residing on chromosomes 2B, 3AL, 3B and 4AL (three out of six environments each). The most significant MTA ($P < 0.001$) was detected on chromosome 2AL for marker *wpt9277* at GRW10. This same marker was consistently associated with days to maturity. Grain filling duration had stable MTA on chromosomes 1BL, 3BS and 7AL (each showing up in three of six environments).

Marker–trait associations were found for plant height on chromosomes 3BL, 5BS, 6AS, 7AS and 7BL, of which the MTAs on chromosomes 6AS and 7BL were the most consistent. QTL regions for flag leaf length were noted on chromosomes 1BS, 1BL, 2BL, 3BL, 3AL and 5B. Most of these associations were consistent, particularly marker *wpt5072* on chromosome 3BL which was detected in five out of six environments. For flag leaf width, however, only w*pt667461* (unmapped) was consistently associated with the trait despite the presence of flag leaf width-associated markers on chromosomes 2DL, 3BL, 5BS, 6A and 7AS.

Table 3 Summary of marker–trait associations detected for agronomic traits and drought-related indices detected in five environments and combined across environments

Trait	Environment[a]							R^2 (%)[b]
	GRW10	GRW11	GRD11	MLKW11	MLKD11	Combined	Total	
BM		1	1	4			6	4.1–5.4
DH	4	3	3	3	3	4	20	2.2–5.4
DM	1	1	2	1		1	6	3.3–4.2
DSI_KN							2	4.2–4.7
DSI_YLD							1	4.2
GA	8	7	8			9	32	2.5–4.0
GFD	9	3	2	3	2	4	23	3.1–6.3
HI		5	1	5	4	2	17	3.0–5.1
SKH	23	30	21	5		34	113	2.8–4.1
KN	1	4	2	1	3	4	15	2.8–5.0
KNL	1		4				5	3.5–4.4
KNS	3		2				5	3.9–4.4
KWS	5	1				5	11	2.3–4.4
LA	6	3	6		3	7	25	1.9–3.8
LL	8	5	4	5	6	11	39	2.6–5.3
LW	3	1	2	1	6	2	15	3.2–5.2
LS	1	2	1				4	4.4.–4.5
NDVI	2	2	1			5	10	2.0–3.6
PHT	7	5	3	8		9	32	2.2–5.3
SKD	6	8	9			15	38	2.0–4.2
SKW	8	8	6			16	38	1.8–4.3
SL	1	4	3				8	2.0–2.3
SN	4	1	6			7	18	3.0–8.2
SPN			4	1			5	3.7–6.9
TKW	1	5	3	3	4	5	21	2.1–5.2
TW	6	8	15			15	44	2.2–4.8
YLD		3	3	4	2	3	15	2.8–5.2

BM biomass, *DM* days to maturity, *DH* days to heading, *DSI_KN* drought susceptibility index calculated from kernel number, *DSI_YLD* drought susceptibility index calculated from grain yield, *GA* leaf green area, *GFD* grain filling duration, *HI* harvest index, *SKH* hardness index, *KN* grain number, *KNL* kernel number per spikelet, *KNS* kernel number per spike, *KWS* kernel weight per spike, *LA* leaf area, *LL* flag leaf length, *LW* flag leaf width, *LS* leaf senescence, *NDVI* normalized difference vegetation index, *PHT* plant height, *SKD* single-kernel diameter, *SKW* single-kernel weight, *SL* spike length, *SN* spikes number per meter square, *SPN* spikelet number, *TKW* thousand-kernel weight, *TW* test weight, *YLD* grain yield

[a] Environment at which MTA detected for each trait. GRW10, Greeley irrigated 2010; GRW11, Greeley Irrigated 2011; GRD11, Greeley rainfed 2011; MLKW11, Melkassa non-stressed 2011; MLKD11, Melkassa stressed 2011; Combined, combined data across environments

[b] Phenotypic variation explained by markers without cross-validation

Both stable and environment-specific MTAs were detected for flag leaf area; the chromosomes 3BL and 5BL harbored stable QTL for this trait. Unmapped markers *wpt0605* and *wpt1370* were also consistently associated with flag leaf area.

Significant MTA were also obtained for drought tolerance-related traits and vegetation indices. Regions of chromosome 4AL, 7A and 7BL comprised QTL for drought susceptibility index. Leaf senescence QTL were found in three regions of chromosome 6B (36.1, 50.6 and 84.6 cM) and another five unmapped markers also showed associations with leaf senescence. Regions of chromosomes 1AL, 1BS, 2AS and 6BL harbored QTL for NDVI, and unmapped marker *wpt0694* was also associated with NDVI in two environments.

Some of the MTA were significant at false discovery rate (FDR = 0.05) after correcting for multiple comparisons. These significant MTA at FDR = 0.05 were obtained for spikelet number per spike on chromosome 2BS, plant height on chromosome 6AS, grain filling duration on chromosome 3BS and green leaf area on chromosome 1BL. Associations of unmapped markers with flag leaf width

Table 4 Marker–trait associations significant at FDR = 0.05 for phenotypic traits measured in five environments

Trait	Environment	Marker	Chromosome	Position (cM)	R^2 (%)[a]	R^2 (%)[b]	FDR *P* value[c]
GA	Combined	*wPt4532*	1BL	88.3	5.6	14.5	0.0097
GA	Combined	*wPt0944*	Unmapped		4.9	14.0	0.0206
LW	MLKD11	*wPt730263*	Unmapped	9.0	7.3	0.9	0.014
PHT	Combined	*wPt729839*	6AS	45.4	5.3	21.8	0.0071
SPN	MLKW11	*wPt8492*	2BS	65.7	7.0	6.1	0.0058
SN	GRW10	*wPt666595*	Unmapped		8.2	2.4	0.0054
SN	GRW10	*wPt667101*	Unmapped		6.6	0.5	0.0260
GFD	GRW10	*tPt9267*	3BS	24.6	6.3	7.4	0.0317
GFD	GRW10	*wPt5836*	3BS	39.1	6.0	3.4	0.0317
GFD	GRW10	*wPt798970*	3BS	25.1	5.6	1.7	0.0353
DSI_KN	Melkassa	*wPt0419*	Unmapped		6.0	1.1	0.0226

GA green leaf area, *LW* leaf width, *PHT* plant height, *SPN* spikelet number per spike, *SN* spike number, *DSI_KN* drought susceptibility index calculated from kernel number

[a] Phenotypic variation explained by markers without cross-validation

[b] Phenotypic variation explained by markers after cross-validation

[c] False discovery rate *P* value

(*wpt730263*), spike number (*wpt666595* and *wpt667101*) and drought susceptibility index (*wpt0419*) were also significant at FDR = 0.05 (Table 4). None of the MTA obtained for the remaining traits survived the FDR adjustment for multiple testing. At a relaxed FDR of 0.25, however, MTAs were identified for test weight, biomass, leaf green area, harvest index, leaf length, leaf width, single-kernel diameter, kernel hardness, flag leaf area, kernel per spikelet and kernel number-based drought susceptibility index (data not shown). The non-cross-validated proportion of phenotypic variances explained by markers significant at FDR = 0.05 were larger than the cross-validated proportions of phenotypic variances in most cases (Table 4).

Multi-trait MTAs were detected in many chromosome regions. Their chromosome positions are shown with other trait-specific QTL in Fig. 5. Clusters of QTL were detected for kernel size-related traits on chromosomes 1BL, 4AL and 7DL. Kernel quality traits (SKH and TW) had QTL in common with one or more kernel size-related traits on chromosomes 1D, 2DL, 3BS, 3D, 4AL, 5B and 7AS. Markers near the centromeric region of chromosome 5B (67.7–76.4 cM) were associated with yield, spike number per square meter, harvest index, biomass, plant height, thousand-kernel weight and test weight. A region of chromosome 1AL was associated with both harvest index and NDVI, and QTL for green leaf area was detected close to the region of harvest index on chromosome 5AL. The QTL on 1BL for leaf green area was in the same region with a QTL detected for SKD, while green leaf area QTL on 3BL was close to the QTL region for TKW, TW and GFD. Similarly, yield and harvest index had QTL in common on chromosome 1BS. Flag leaf area and flag leaf length had QTL in common on chromosomes 3BL and 5BL.

Multi-trait QTL were also detected among drought tolerance-related traits and vegetation indices. Marker–trait associations were obtained on chromosome 6BL for NDVI, leaf senescence and leaf green area index. However, only MTA for leaf senescence and leaf green area index were roughly in the same region (within 3 cM). Among QTL detected for drought susceptibility index, QTL on chromosome 4AL was in the same region with the QTL for single-kernel diameter, single-kernel weight and thousand-kernel weight. Similarly, a QTL on 7A was detected in the same region with QTL detected for kernel number, kernel number per spikelet and kernel number per spike. Drought susceptibility QTL on 7BL was detected at a distance of 3.4 and 3.9 cM away from plant height and spike number QTL, respectively.

Discussion

The spring wheat association mapping population panel (WAMII) used in this study was developed by CIMMYT with the intention of identifying QTL/genes underlying drought and heat tolerance-related traits. While the panel was assembled deliberately to restrict the range of phenology which otherwise can mask detection of QTL (Reynolds et al. 2009; Pinto et al. 2010), accessions in the panel had wide differences in morphological characters and agronomic traits. These allowed us to apply a genome-wide association mapping approach for studying the genetic basis of phenotypic variation for traits evaluated under a

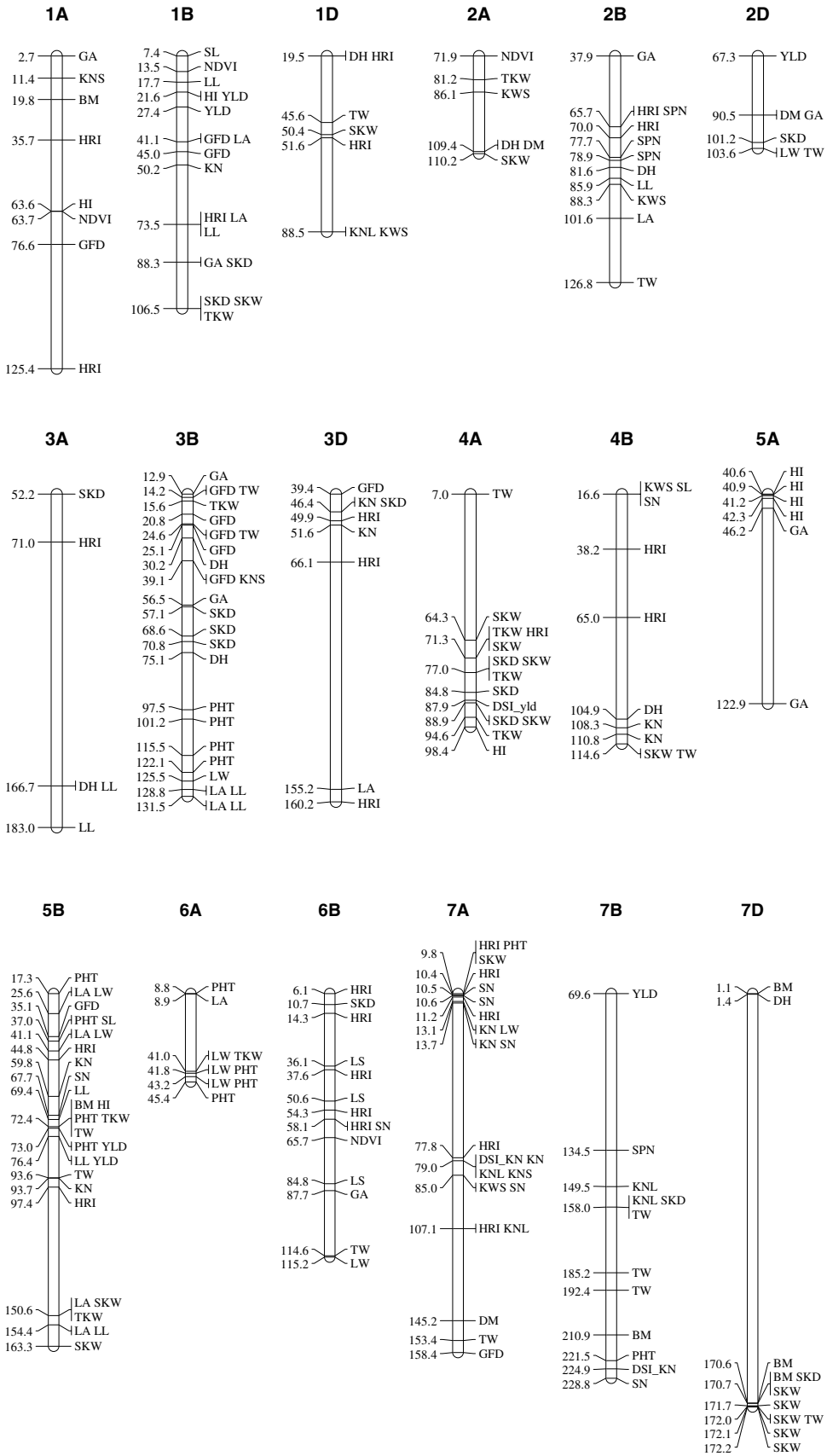

Fig. 5 Chromosomal regions of significant marker–trait associations identified for phenotypic traits measured in this study. The DArT marker positions are provided in Online Resource 4

wide range of environmental factors. The accessions have been exposed to water stress (as low as 192 mm for the entire growing season under rain-fed conditions at Greeley in 2011) and heat stress (maximum temperature >30 °C for a majority of the days after heading throughout the grain filling period at Greeley). For wheat anthesis and grain filling, the optimum temperature ranges from 12 to 22 °C (Farooq et al. 2011). Temperatures above 30 °C during floret formation in wheat may lead to complete sterility (Saini and Aspinall 1982).

Population structure can lead to false associations between markers and traits if not taken into account during association analysis (Zhao et al. 2007). A model-based approach was used to detect subgroups for 287 spring wheat lines in the association mapping panel, and we were able to detect seven subpopulations. In the molecular variance analysis of our study, the significant ($P < 0.001$) population differentiation (Fst ranged from 0.14 to 0.73) for the seven groups reaffirms the presence of population structure. Genetic distance-based cluster analysis also provided evidence for the presence of subpopulations despite the lack of similarity between its clusters and the subgroups of model-based analysis in STRUCTURE. The majority of the variation was explained by within-population variation (78.5 %), with among-population variation accounting for 21.5 % of the variation. The higher within-population variation demonstrates the impact of selection in maintaining allele diversity in the breeding populations. The magnitude of among-population variation in this study is comparable to variation explained due to differences between European and Asian wheat germplasm (Hao et al. 2010), and even higher than the variation explained due to differences among geographical groups of wheat populations in Europe (Roussel et al. 2005). Chao et al. (2010) reported a higher among-subpopulation genetic variation in spring wheat (17.2 %) than in winter wheat (10.5 %) from the United States and CIMMYT breeding programs. Although lines from CIMMYT were used in the present study, the subpopulations are more genetically differentiated than the US spring wheat subpopulations in the study by Chao et al. (2010). In our mapping panel, a substantial number of lines shared one or more parents. Therefore, some of the groups were dominated by lines that trace back to a common parent (data not shown). Because a few elite lines are routinely used as parents of crosses in many breeding programs, this can be expected to lead to some sort of population structure as observed in the current study.

Linkage disequilibrium information is critical in association studies because LD values can be affected by many factors such as population type, chromosome region and mating system. The number of markers needed for association studies depends on the extent of LD under consideration. In the current population, chromosomes showed large differences in the proportion of marker pairs in significant LD ($P < 0.01$) from the maximum 62 % for chromosome 4A to the minimum 20 % for chromosome 5A. Although chromosome 4A contained more markers in LD at $P < 0.01$, markers on chromosome 3D are more physically linked ($r^2 = 0.264$). However, the proportions of marker pairs at $r^2 = 0.2$ and $r^2 = 0.264$ are comparable for all chromosomes implying the importance of choosing an appropriate r^2 value as a threshold in addition to statistical significance. In the current analysis, $r^2 = 0.2$ was used only for chromosomes with weak LD which do not allow evaluating LD decay rate at threshold level of physically linked markers ($r^2 = 0.264$).

The magnitude of LD across a genome or chromosome is a function of nucleotide or linkage distance. LD decay rate was determined both at the genome and individual chromosome level. LD decayed within 2 cM for both A and B genomes, while it extended up to ~6.8 cM of genetic distance for the D genome. Chao et al. (2010) reported a similar finding using 394 genetically mapped SNP markers on 478 spring and winter wheat cultivars. The reason for more extended LD in the D-genome than in the A- and B-genomes could be the introduction of new haplotypes, which can increase the extent of LD from *Aegilops tauschii* (D-genome donor) into the D-genome of hexaploid wheat germplasm through synthetic wheats. Many lines with synthetic background have been included during assembly of this association mapping panel (Lopes et al. 2012). Another potential explanation for extended LD is the genetic bottleneck that occurred on the D-genome as a result of hybridization of tetraploid wheat with *Ae. tauschii* to form hexaploid wheat (Warburton et al. 2006; Chao et al. 2010).

The LD decay rate was also determined for subpopulations. Generally, LD extended over longer nucleotide or genetic distances (6–9 cM) for subpopulations than the whole panel, which is expected because grouping of genetically similar genotypes reduces within subpopulation genetic diversity; consequently, large blocks of a chromosome region could be in LD. We were able to fit LOESS curves only for four out of the seven model-based subpopulations; three of them had similar LD decay rates (within 8–9 cM), and for one group LD decayed relatively faster (within 6 cM) at $r^2 = 0.2$. When all subpopulations were considered together, on average LD decayed below the base line $r^2 = 0.2$ at ~3.4 cM which is about 50 % of the genetic distance within subpopulations. In other words, this translated to the doubling of genetic distance over which LD extended just by grouping similar genotypes together or using closely related genotypes for assessing LD level.

The effect of subpopulations on phenotypic traits was assessed with multiple regression analysis. Among the traits confounded by population structure, plant height, kernel traits (single-kernel weight, thousand-kernel weight

and single-kernel diameter) and NDVI showed the greatest percentage of phenotypic variation explained by population structure. Interestingly grain yield, kernel number per square meter, spike number per square meter, harvest index and phenological traits (DH, DM and GFD) were among the group of traits least affected by subpopulations. Except for TKW and HI, these results are in agreement with Dodig et al. (2012) who reported large effects of population structure in winter wheat on stem-related traits (stem height, peduncle length and peduncle extrusion); a moderate influence on sterile spikelet per spike and biomass per plant; and a low effect on yield and yield components (KN, TKW and SN). The greatest effect of population structure on kernel-size related traits in our panel may be due to intensive selection for kernel size in CIMMYT's breeding program (Ravi Singh, CIMMYT, personal comm.). Elite lines are most likely larger in kernel size than the remaining lines included in the panel for the purpose of maintaining genetic diversity during assembly of the mapping panel. The low effect of population structure on heading date indicates the minimum confounding effect of phenology on population structure, unlike plant height and kernel size.

Although grain yield QTL were detected on all wheat chromosomes in previous studies, relatively consistent MTA in our study were detected on chromosomes 1BS, 2DS, 5B and 7B. Broad comparison of MTA results from the current study with previous studies were made using chromosome arms because of differences in marker type and marker positions on different genetic maps. The DArT marker *wpt6531* on the short arm of chromosome 2DS, which was associated with yield in the current study, is about 8 cM away from the *wpt4144* marker, which was associated with yield in the Crossa et al. (2007) study. Kumar et al. (2007) detected QTL for yield in this region linked to SSR marker *gwm261* which is 14.4 cM distal to *Ppd-D1* and 0.6 cM distal to the height-reducing semi-dwarfing *Rht8* gene (Korzun et al. 1998; Ellis et al. 2007). Dodig et al. (2012) also detected QTL on chromosome 2DS (near *gwm484*) that explained about 22 % of the phenotypic variation for grain yield. Therefore, the stable and highly significant grain yield MTA on 2DS in the current study is probably due to a grain yield QTL in proximity to the *Ppd-D1* locus, which is known for its influence on wheat yield through optimization of flowering time (Worland 1996). Significant MTA for yield were detected on the short arm of chromosome 1B in the Crossa et al. (2007) study. Quarrie et al. (2005) found major QTL which explained up to 35 % of the phenotypic variation and were expressed in 11 out of 24 trials on 7BL.

In the current study, *wpt8211* on chromosome 7B (69.6 cM) was associated with yield in three environments. The marker *wpt3457* on chromosome 5B (73 cM) was associated with yield both under irrigated and rain-fed conditions, and another marker (*wpt6135*), which was physically in LD and 3.4 cM away from *wpt3457*, was strongly associated with yield under irrigated conditions. Moreover, many other markers consistently associated with traits such as thousand-kernel weight, final biomass, harvest index, plant height and flag leaf length also reside on either side of the yield QTL position on 5B, indicating the importance of this region in influencing yield and yield components. This region may explain a portion of the genotypic correlations of yield with yield component traits.

In previous studies, yield QTL have been detected on both long and short arms of chromosome 5B (Huang et al. 2003; Marza et al. 2006; Crossa et al. 2007; Neumann et al. 2011) and some of those QTL may coincide with the QTL detected here on chromosome 5B. However, to our knowledge, there are no reports on the presence of multi-trait QTL near the centromeric region of chromosome 5B. In fact, chromosome 5B comprised the highest number of MTA in this study. Kumar et al. (2007) reported multi-trait QTL for yield and yield components on chromosomes 2DS and 4AL. However, no multi-trait regions were detected on 2DS for the yield component traits in this study, but a region of chromosome 4AL was identified as a multi-trait region for kernel size and quality traits. Similar results were observed by Lopes et al. (2013).

Grain yield and harvest index shared an association region on chromosome 1BS, implying that there is a genetic basis for the high and consistent genotypic correlation observed between grain yield and harvest index (data not shown). In addition, NDVI has QTL in common only with harvest index on chromosome 1AL of all yield component traits, while green leaf area shared QTL with harvest index on chromosomes 5A (42.3 cM) and with single-kernel diameter on chromosome 3B (56.5 cM). The benefits of assessing yielding ability of wheat with these vegetation indices may be dictated by the expression of genes in the chromosome regions that harvest index and single-kernel diameter shared with the indices.

Trait-specific stable MTA were detected for main yield component traits such as kernel number per square meter, harvest index and thousand-kernel weight. Unlike harvest index and thousand-kernel weight, only one marker on chromosome 7AS showed consistency across environments for kernel number. Among environment-specific MTA for kernel number, the unmapped marker *wpt0866* had sequence similarity with 1, 3-beta glucan synthase (Marone et al. 2012). Similarly, all MTA of final biomass, kernel number per spikelet and kernel number per spike were environment specific, showing the presence of higher genotype by environment interaction for these yield component traits than yield itself. Among yield and yield component traits, however, very strong (FDR = 0.05) MTAs were obtained for spikelet number per spike on chromosome

2BS (*wpt8492*), and for spike number per square meter for two unmapped markers *wpt666595* and *wpt667101*. None of these MTAs are in agreement with previously detected QTL in bi-parental populations with the exception of MTA noted for spikelet number on chromosomes 2B and 7B by Neumann et al. (2011) which may be comparable with our current findings.

Stacking QTL that control traits of interest from different chromosome regions into one background is a challenging and time consuming task in plant breeding. Using multi-trait markers in marker-assisted selection may increase QTL pyramiding efficiency. With the exception of chromosomes 4D, 5D and 6D, two or more traits shared the same region or resided within 5 cM on all chromosomes. Kernel size-related traits, single-kernel weight, single-kernel diameter and thousand-kernel weight had QTL in common on chromosomes 1BL, 4AL (SKW, SKD and TKW) and 7DL (SKW and SKD). Test weight also shared the same regions with one or more kernel size-related traits on chromosomes 1B, 2DL, 4BL, 7BL and 7DL. These traits could be under the same genetic control and markers in those multi-trait regions could be used in the future for improvement of kernel size-related traits through marker-assisted selection. Similarly, clusters of QTL for flag leaf characters (LA, LL and LW) were found on chromosomes 3BL and 5BL. Moreover, there is a pattern of co-localization of QTL for leaf characters and kernel size-related traits TKW or SKW. This may be related to the translocation of flag leaf photosynthetic products to growing kernels during the grain filling period (Lupton 1966).

Many chromosome regions were associated with drought tolerance-related traits such as drought susceptibility index, NDVI, leaf senescence, green leaf area and flag leaf characters. Overall, the most important chromosomes that comprised QTL for drought tolerance in this study are chromosomes 1B, 4AL, 6B, 5B, 7A and 7B. The major drought tolerance QTL detected in the past correspond to chromosome regions associated with key drought tolerance-related traits in this study (Alexander et al. 2012) at least at chromosome arms level.

Although there was a wide range in mean phenotypic values for plant height, the major plant height reducing genes *Rht-D1b* and *Rht-B1b* were not detected in this study. However, we detected plant height MTA in the regions of previously reported QTL on chromosomes 3BL (Maccaferri et al. 2011), 5B (Cadalen et al. 1997; McCartney et al. 2005; McIntyre et al. 2010), 6AS (Spiel Meyer et al. 2005) and 7BL (McCartney et al. 2005). Similarly, regions of group 5 chromosomes where *VRN-1* genes reside were not associated with heading date in this panel. Nonetheless, in agreement with the result in this study, QTL that affect flowering time in wheat have been reported on chromosomes 2B, 3AL, 3B and 7DS (Borner et al. 2002; Marza et al. 2006; Cuthbert et al. 2008; Wang et al. 2009).

Photoperiod sensitivity genes, which have been mapped on the short arms of homeologous group 2 chromosomes, were not detected for heading date. However, the QTL detected on 3AL may indicate variation in an earliness per se gene, which was mapped on chromosome 3AL (Borner et al. 2002).

In conclusion, we have shown that LD decay varied both at the genome and chromosome levels. Genome-wide association mapping effectively detected both stable and environment-specific QTL for yield, yield components, and drought-related traits. Multi-trait chromosome regions have been detected and particularly the region on chromosome 5B associated with yield and yield component traits may be useful in MAS following proper validation. In the context of drought tolerance, QTL regions that control both drought tolerance-related traits and yield component traits were detected on chromosomes 1AL (NDVI and harvest index), 5AL (green leaf area and harvest index) and 3B (green leaf area and single-kernel diameter), implying the possibility of using vegetation indices for indirect assessment for certain yield component traits.

Acknowledgments This research was conducted with the financial support from Beachell-Borlaug International Scholarship award to the first author.

Conflict of interest The authors declare that they have no conflict of interest.

Ethical standards This experiment complies with the current laws of the US and Ethiopia where the experiment was performed.

References

Abdurakhmonov I, Abdukarimov A (2008) Application of association mapping to understanding the genetic diversity of plant germplasm resources. In J Plant Genom. doi:10.1155/2008/574927

Adhikari TB, Gurung S, Hansen JM, Jackson EW, Bonman JM (2012) Association mapping of quantitative trait loci in spring wheat landraces conferring resistance to bacterial leaf streak and spot blotch. Plant Genome 5:1–16. doi:10.3835/plantgenome2011.12.0032

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang SY, Uszynski G, Mohler V, Lehmensiek A,Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A(2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploidwheat genome. Theor Appl Genet 113:1409–1420. doi:10.1007/s00122-006-0365-4

Alexander LM, Kirigwi FM, Fritz AK, Fellers JP (2012) Loci analysis of drought tolerance in a spring wheat population using amplified fragment length polymorphism and diversity array technology markers. Crop Sci 52:253–261

Araus JL, Slafer GA, Royo C, Serret MD (2008) Breeding for yield potential and stress adaptation in cereals. Crit Rev Plant Sci 27:377–412. doi:905611485

Bagge M, Xia XC, Lubberstedt T (2007) Functional markers in wheat—commentary. Curr Opin Plant Biol 10:211–216. doi:101016/jpbi200701009

Barnabas B, Jager K, Feher A (2008) The effect of drought and heat stress on reproductive processes in cereals. Plant Cell Environ 31:11–38. doi:101111/j1365-3040200701727x

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B 57(1):289–300

Benson J, Brown-Guedira G, Murphy JP, Sneller C (2012) Population structure, linkage disequilibrium, and genetic diversity in soft winter wheat enriched for fusarium head blight resistance. Plant Genome 5:71–80. doi:103835/plantgenome2011110027

Borner A, Schumann E, Furste A, Coster H, Leithold B, Roder MS, Weber WE (2002) Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). Theor Appl Genet 105:921–936. doi:101007/s00122-002-0994-1

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:101093/bioinformatics/btm308

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. Genetics 172:1165–1177. doi:101534/genetics105044586

Cadalen T, Boeuf C, Bernard S, Bernard M (1997) An intervarietal molecular marker map in *Triticum aestivum* L em Thell and comparison with a map from a wide cross. Theor Appl Genet 94:367–377. doi:101007/s001220050425

Casadesus J, Kaya Y, Bort J, Nachit MM, Araus JL, Amor S, Ferrazzano G, Maalouf F, Maccaferri M, Martos V, Ouabbou H, Villegas D (2007) Using vegetation indices derived from conventional digital cameras as selection criteria for wheat breeding in water-limited environments. Ann Appl Biol 150:227–236. doi:101111/j1744-7348200700116x

Chao SM, Zhang WJ, Dubcovsky J, Sorrells M (2007) Evaluation of genetic diversity and genome-wide linkage disequilibrium among US wheat (*Triticum aestivum* L.) germplasm representing different market classes. Crop Sci 47:1018–1030. doi:102135/cropsci2006060434

Chao SM, Dubcovsky J, Dvorak J, Luo MC, Baenziger SP, Matnyazov R, Clark DR, Talbert LE, Anderson JA, Dreisigacker S, Glover K, Chen JL, Campbell K, Bruckner PL, Rudd JC, Haley S, Carver BF, Perry S, Sorrells ME, Akhunov ED (2010) Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). BMC Genom 11:727 doi:10.1186/1471-2164-11-727 (Artn 727)

Cleveland WS (1979) Robust locally-weighted regression and smoothing scatterplots. J Am Stat Assoc 74:829–836

Collins NC, Shirley NJ, Saeed M, Pallotta M, Gustafson JP (2008) An ALMT1 gene cluster controlling aluminum tolerance at the Alt4 locus of rye (*Secale cereale* L.). Genetics 179:669–682. doi:101534/genetics107083451

Comadran J, Thomas WTB, van Eeuwijk FA, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, Romagosa I, Hackett CA, Russell JR (2009) Patterns of genetic diversity and linkage disequilibrium in a highly structured *Hordeum vulgare* association-mapping population for the Mediterranean basin. Theor Appl Genet 119:175–187. doi:101007/s00122-009-1027-0

Cooper M, van Eeuwijk FA, Hammer GL, Podlich DW, Messina C (2009) Modeling QTL for complex traits: detection and context for plant breeding. Curr Opin Plant Biol 12:231–240. doi:101016/jpbi200901006

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J, Reynolds M, Crouch JH, Ortiz R (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177:1889–1913. doi:101534/genetics107078659

Cuthbert JL, Somers DJ, Brule-Babel AL, Brown PD, Crow GH (2008) Molecular mapping of quantitative trait loci for yield and yield components in spring wheat (*Triticum aestivum* L.). Theor Appl Genet 117:595–608. doi:10.1007/s00122-008-0804-5

Dodig D, Zoric M, Kobiljski B, Savic J, Kandic V, Quarrie S, Barnes J (2012) Genetic and association mapping study of wheat gronomic traits under contrasting water Regimes. Int J Mol Sci 13:6167–6188. doi:103390/Ijms13056167

Ellis M, Bonnett D, Rebetzke G (2007) A 192 bp allele at the Xgwm261 locus is not always associated with the Rht8 dwarfing gene in wheat (*Triticum aestivum* L.). Euphytica 157:209–214

Ersoz E, Yu J, Buckler E (2009) Applications of linkage disequilibrium and association mapping in maize. In: Kriz AL, Larkins BA (eds) Molecular genetic approaches to maize improvement 173, vol. 63. Biotechnology in Agriculture and Forestry

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620. doi:101111/j1365-294X200502553x

Farooq M, Bramley H, Palta J, Siddique K (2011) Heat stress in wheat during reproductive and grain-filling phases. Crit Rev Plant Sci 30:491–507

Fisher RA, Maurer R (1978) Drought resistance in spring wheat cultivars: I. Grain yield responses. Aust J Agric Res 29:897–912

Fleury D, Jefferies S, Kuchel H, Langridge P (2010) Genetic and genomic tools to improve drought tolerance in wheat. J Exp Bot 61:3211–3222. doi:101093/Jxb/Erq152

Glover KD, Rudd JC, Devkota RN, Hall RG, Jin Y, Rickertsen JR (2006) Registration of 'Granger' wheat. Crop Sci 46:1390–1391

Green AJ, Berger G, Griffey CA, Pitman R, Thomason W, Balota M, Ahmed A (2012) Genetic yield improvement in soft red winter wheat in the Eastern United States from 1919 to 2009. Crop Sci 52:2097–2108. doi:102135/cropsci2012010026

Gupta PK, Varshney RK, Sharma PC, Ramesh B (1999) Molecular markers and their applications in wheat breeding. Plant Breed 118:369–390. doi:101046/j1439-0523199900401x

Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. Plant Mol Biol 57:461–485. doi:101007/s11103-005-0257-z

Habash DZ, Kehel Z, Nachit M (2009) Genomic approaches for designing durum wheat ready for climate change with a focus on drought. J Exp Bot 60:2805–2815. doi:101093/Jxb/Erp211

Hao CY, Perretant MR, Choulet F, Wang LF, Paux E, Sourdille P, Zhang XY, Feuillet C, Balfourier F (2010) Genetic diversity and linkage disequilibrium studies on a 3.1-Mb genomic region of chromosome 3B in European and Asian bread wheat (*Triticum aestivum* L.) populations. Theor Appl Genet 121:1209–1225. doi:101007/s00122-010-1382-x

Holland JB (2007) Genetic architecture of complex traits in plants. Curr Opin Plant Biol 10:156–161. doi:101016/jpbi200701003

Huang XQ, Coster H, Ganal MW, Roder MS (2003) Advanced backcross QTL analysis for theidentification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). Theor Appl Genet 106:1379–1389. doi: 10.1007/s00122-002-1179-7

Kollers S, Rodemann B, Ling J, Korzun V, Ebmeyer E, Argillier O, Hinze M, Plieske J, Kulosa D, Ganal MW, Röder MS (2013a) Whole genome association mapping of Fusarium head blight resistance in European winter wheat (*Triticum aestivum* L.). PLoS ONE 8:e57500

Kollers S, Rodemann B, Ling J, Korzun V, Ebmeyer E, Argillier O, Hinze M, Plieske J, Kulosa D, Ganal MW, Röder MS (2013b)

Genetic architecture of resistance to Septoria tritici blotch (*Mycosphaerella graminicola*) in European winter wheat. Mol Breed 32:411–423

Korzun V, Roder M, Ganal M, Worland A, Law C (1998) Genetic analysis of the dwarfing gene (Rht8) in wheat. Part I. Molecular mapping of Rht8 on the short arm of chromosome 2D of bread wheat (*Triticum aestivum* L.). Theor Appl Genet 96:1104–1109

Kumar N, Kulwal PL, Balyan HS, Gupta PK (2007) QTL mapping for yield and yield contributing traits in two mapping populations of bread wheat. Mol Breed 19:163–177. doi:101007/s11032-006-9056-8

Lipka AE, Tian F, Wang QS, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang ZW (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399. doi:101093/bioinformatics/bts444

Lopes MS, Reynolds MP (2012) Stay-green in spring wheat can be determined by spectral reflectance measurements (normalized difference vegetation index) independently from phenology. J Exp Bot 63:3789–3798. doi:101093/Jxb/Ers071

Lopes MS, Reynolds MP, Jalal-Kamali MR, Moussa M, Feltaous Y, Tahir ISA, Barma N, Vargas M, Mannes Y, Baum M (2012) The yield correlations of selectable physiological traits in a population of advanced spring wheat lines grown in warm and drought environments. Field Crops Res 128:129–136. doi:101016/j fcr201112017

Lopes MS, Reynolds MP, McIntyre CL, Mathews KL, Kamali MRJ, Mossad M, Feltaous Y, Tahir ISA, Chatrath R, Ogbonnaya F, Baum M (2013) QTL for yield and associated traits in the Seri/Babax population grown across several environments in Mexico, in the West Asia, North Africa, and South Asia regions. Theor Appl Genet 126:971–984. doi:101007/s00122-012-2030-4

Lupton FH (1966) Translocation of photosynthetic assimilates in wheat. Ann Appl Biol 57:355-364. doi: 10.1111/j.1744-7348.1966.tb03829.x

Maccaferri M, Sanguineti MC, Mantovani P, Demontis A, Massi A, Ammar K, Kolmer JA, Czembor JH, Ezrati S, Tuberosa R (2010) Association mapping of leaf rust response in durum wheat. Mol Breed 26:189–228. doi:101007/s11032-009-9353-0

Maccaferri M, Sanguineti MC, Demontis A, El-Ahmed A, del Moral LG, Maalouf F, Nachit M, Nserallah N, Ouabbou H, Rhouma S, Royo C, Villegas D, Tuberosa R (2011) Association mapping in durum wheat grown across a broad range of water regimes. J Exp Bot 62:409–438. doi:101093/Jxb/Erq287

Marone D, Panio G, Ficco DBM, Russo MA, De Vita P, Papa R, Rubiales D, Cattivelli L, Mastrangelo AM (2012) Characterization of wheat DArT markers: genetic and functional features. Mol Genet Genom 287:741–753. doi:101007/s00438-012-0714-8

Marza F, Bai GH, Carver BF, Zhou WC (2006) Quantitative trait loci for yield and related traits in the wheat population Ning7840 × Clark. Theor Appl Genet 112:688–698. doi: 10.1007/s00122-005-0172-3

McCartney CA, Somers DJ, Humphreys DG, Lukow O, Ames N, Noll J, Cloutier S, McCallum BD (2005) Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452 x 'AC Domain'. Genome 48:870–883

McIntyre CL, Mathews KL, Rattey A, Chapman SC, Drenth J, Ghaderi M, Reynolds M, Shorter R (2010) Molecular detection of genomic regions associated with grain yield and yield-related components in an elite bread wheat cross evaluated under irrigated and rainfed conditions. Theor Appl Genet 120:527–541. doi:101007/s00122-009-1173-4

Mergoum M, Frohberg RC, Ali S, Singh PK, Rasmussen JB, Miller JD (2006) Registration of spring wheat germplasm ND 735 combining tan spot, leaf, and stem rusts. Crop Sci 46:1003–1004

Mir R, Zaman-Allah M, Sreenivasulu N, Trethowan R, Varshney R (2012) Integrated genomics, physiology and breeding approaches for improving drought tolerance in crops. Theor Appl Genet. doi:10.1007/s00122-012-1904-9

Neumann K, Kobiljski B, Dencic S, Varshney RK, Borner A (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). Mol Breed 27:37–58. doi:101007/s11032-010-9411-7

Passioura JB (2012) Phenotyping for drought tolerance in grain crops: when is it useful to breeders? Funct Plant Biol 39:851–859. doi:101071/Fp12079

Peng JH, Bai Y, Haley SD, Lapitan NLV (2009) Microsatellite-based molecular diversity of bread wheat germplasm and association mapping of wheat resistance to the Russian wheat aphid. Genetica 135:95–122. doi:101007/s10709-008-9262-x

Pinto RS, Reynolds MP, Mathews KL, McIntyre CL, Olivares-Villegas JJ, Chapman SC (2010) Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. Theor Appl Genet 121:1001–1021. doi:101007/s00122-010-1351-4

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

Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C, Steele N, Pljevljakusic D, Waterman E, Weyen J, Schondelmaier J, Habash DZ, Farmer P, Saker L, Clarkson DT, Abugalieva A, Yessimbekova M, Turuspekov Y, Abugalieva S, Tuberosa R, Sanguineti MC, Hollington PA, Aragues R, Royo A, Dodig D (2005) A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring X SQ1 and its use to compare QTLs for grain yield across a range of environments. Theor Appl Genet 110:865–880. doi:101007/s00122-004-1902-7

Ravi K, Vadez V, Isobe S, Mir RR, Guo Y, Nigam SN, Gowda MVC, Radhakrishnan T, Bertioli DJ, Knapp SJ, Varshney RK (2011) Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance in groundnut (*Arachis hypogaea* L.). Theor Appl Genet 122:1119–1132

Reynolds MP, Manes Y, Izanloo A, Langridge P (2009) Phenotyping for physiological breeding and gene discovery in wheat. Ann Appl Biol 155:309–320

Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F (2005) SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor Appl Genet 111:162–170. doi:101007/s00122-005-2014-8

Saini HS, Aspinall D (1982) Abnormal sporogenesis in wheat (*Triticum aestivum* L.) induced by short periods of high temperature. Ann Bot 49:835–846

Shen X (2013) Revealing the missing heritability via cross-validated genome wide association studies. Published online at http://arxiv.org/ftp/arxiv/papers/1307/1307.7950.pdf

Sinclair TR (2012) Is transpiration efficiency a viable plant trait in breeding for crop improvement? Funct Plant Biol 39:359–365. doi:101071/Fp11198

Sorrells M, Yu J (2009) Linkage disequilibrium and association mapping in the Triticeae. In: Feuillet C, Muehlbauer GJ (eds) Genetics and genomics of the triticeae, plant genetics and genomics: crops and models 7:655–683. doi:10.1007/978-387-77489-3_22

Spiel Meyer W, McIntosh RA, Kolmer J, Lagudah ES (2005) Powdery mildew resistance and Lr34/Yr18 genes for durable resistance to leaf and stripe rust cosegregate at a locus on the short arm of chromosome 7D of wheat. Theor Appl Genet 111:731–735. doi:101007/s00122-005-2058-9

Wang RX, Hai L, Zhang XY, You GX, Yan CS, Xiao SH (2009) QTL mapping for grain filling rate and yield-related traits in RILs of the Chinese winter wheat population Heshangmai × Yu8679. Theor Appl Genet 118:313–325. doi: 10.1007/s00122-008-0901-5

Warburton ML, Crossa J, Franco J, Kazi M, Trethowan R, Rajaram S, Pfeiffer W, Zhang P, Dreisigacker S, van Ginkel M (2006)

Bringing wild relatives back into the family: recovering genetic diversity in CIMMYT improved wheat germplasm. Euphytica 149:289–301. doi:10.1007/s10681-005-9077-0

Waugh R, Jannink JL, Muehlbauer GJ, Ramsay L (2009) The emergence of whole genome association scans in barley. Curr Opin Plant Biol 12:218–222. doi:101016/jpbi200812007

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. Plant Cell 16:2719–2733

Worland AJ (1996) The influence of flowering time genes on environmental adaptability in European wheats. Euphytica 89:49–57

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208. doi:101038/Ng1702

Yu LX, Lorenz A, Rutkoski J, Singh RP, Bhavani S, Huerta-Espino J, Sorrells ME (2011) Association mapping and gene–gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theor Appl Genet 123:1257–1268. doi:101007/s00122-011-1664-y

Yu LX, Liu SX, Anderson JA, Singh RP, Jin Y, Dubcovsky J, Brown-Guidera G, Bhavani S, Morgounov A, He ZH, Huerta-Espino J, Sorrells ME (2012) Haplotype diversity of stem rust resistance loci in uncharacterized wheat lines. Mol Breed 30:613–614. doi:101007/s11032-011-9680-9

Zhao HH, Fernando RL, Dekkers JCM (2007) Effects of population structure on power and precision of regression-based linkage disequilibrium mapping of QTL. J Anim Sci 85:63

Zheng S, Byrne PF, Bai GH, Shan XY, Reid SD, Haley SD, Seabourn BW (2009) Association analysis reveals effects of wheat glutenin alleles and rye translocations on dough-mixing properties. J Cereal Sci 50:283–290. doi:10.1016/j.jcs.2009.06.008