ORIGINAL PAPER

Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments

**Sivakumar Sukumaran · Susanne Dreisigacker ·
Marta Lopes · Perla Chavez · Matthew P. Reynolds**

Abstract

Key message **Through genome-wide association study, loci for grain yield and yield components were identified in chromosomes 5A and 6A in spring wheat** ***(Triticum aestivum)***.

Abstract Genome-wide association study (GWAS) was conducted for grain yield (YLD) and yield components on a wheat association mapping initiative (WAMI) population of 287 elite, spring wheat lines grown under temperate irrigated high-yield potential condition in Ciudad Obregón, Mexico, during four crop cycles (from 2009–2010 to 2012–2013). The population was genotyped with high-density Illumina iSelect 90K single nucleotide polymorphisms (SNPs) assay. An analysis of traits across subpopulations indicated that lines with 1B/1R translocation had higher YLD, grain weight, and taller plants than lines without the translocation. GWAS using 18,704 SNPs identified 31 loci that explained 5–14 % of the variation in individual traits. We identified SNPs in chromosome 5A and 6A that were significantly associated with yield and yield components. Four loci were detected for YLD in chromosomes 3B, 5A, 5B, and 6A and the locus in 5A explained 5 % of the variation for grain number/m². The locus for YLD in chromosome 6A also explained 6 % of the variation in grain weight. Loci significantly associated with maturity were identified in chromosomes 2B, 3B, 4B, 4D, and 6A and for plant height in 1A and 6A. Loci were also detected for canopy temperature at grain filling (2D, 4D, 6A), chlorophyll index at grain filling (3B and 6A), biomass (3D and 6A) and harvest index (1D, 1B, and 3B) that explained 5–10 % variation. These markers will be further validated.

Abbreviations

GWAS   Genome-wide association study
SNP   Single nucleotide polymorphism
LD   Linkage disequilibrium

Introduction

Wheat is among the most important food crops and is one of the most traded commodities in the world markets (Curtis and Halford 2014). Genetic gains in the yields of spring wheat in favorable environments averaged 0.6 % annually between 1995 and 2010, based on data from hundreds of testing sites world wide mainly through conventional breeding (Sharma et al. 2012). However, to meet predicted global demand, gains in yield of ~2 % annually, a cumulative increase of 50 % in ~20 years, are required (Lopes et al. 2012b)—a target that would take some 70 years to reach, at a 0.6 % yearly rate of genetic gain. Clearly there is a need to complement conventional breeding approaches with molecular approaches and a better understanding of the genetic basis of yield is a pre-requisite for this.

Genome-wide association studies (GWAS) and quantitative trait loci (QTL) mapping are the two main approaches followed to dissect the genetic bases of complex traits

S. Sukumaran (✉) · S. Dreisigacker · P. Chavez · M. P. Reynolds
Global Wheat Program, International Maize and Wheat
Improvement Center (CIMMYT Int.), Apdo. Postal 6-641,
06600 Mexico, DF, Mexico
e-mail: s.sukumaran@cgiar.org

M. Lopes
CIMMYT, PO Box 39, Emek, Ankara 06511, Turkey

Table 1 Weather data of the experimental site during the crop season in Cd. Obregón, Mexico from 2010 to 2013

Years	Planting date	Seeding emergence	Average MaxT	Average MinT	MaxT > 35 °C	Total rainfall (cm)
2009–2010	20-November-2009	28-November-2009	27.60	10.40	19	42.00
2010–2011	12-November-2010	21-November-2010	28.39	8.60	22	2.00
2011–2012	23-November-2011	30-November-2011	27.94	8.70	23	24.00
2012–2013	21-November-2012	02-December-2012	28.28	9.87	20	7.00

MaxT maximum temperature, *MinT* minimum temperature, *MaxT >35 °C* number of days where MaxT was above 35 °C

(Risch and Merikangas 1996). In wheat, the traditional QTL mapping approach might locate genomic regions with low resolution that are limited to the bi-parental population under study. As a complement to QTL mapping, GWAS offers a high-resolution, cost-effective way for gene discovery and molecular marker identification. Bi-parental populations typically are formed for specific traits, whereas in GWAS, populations are phenotyped for different traits and genotyped once, according to the genetic diversity of the traits in the population (Zhu et al. 2008).

Genome-wide association mapping studies for spring wheat are limited; candidate gene approaches are more common (Chao et al. 2010; Edae et al. 2013). GWAS in wheat is a challenge due the crop's large genome, incomplete genome sequence, and polyploidy, which makes it difficult to assign the markers to individual (A, B, and D) genomes (Sukumaran and Yu 2014). There are a few number of studies that used association mapping to dissect genetic bases of traits in winter wheat (Breseghello and Sorrells 2006; Chao et al. 2010; Poland et al. 2011; Yu et al. 2011; Würschum et al. 2013).

For GWAS, CIMMYT developed a wheat association mapping initiative (WAMI) population that is controlled for phenology and plant height (PH) (Lopes et al. 2012a). Candidate gene association mapping was carried out in this population for drought tolerance using five candidate genes (*DREA1A*, *ERA-1B*, *ERA-1D*, *1-FEH-A*, and *1-FEH-B*) (Edae et al. 2013). The value of the population was initially tested by association analysis with known genes like *Vrn-A1*, *Vrn-B1*, *Vrn-D1*, *Rht-B1*, and *Rht-D1* in chromosomes 5A, 5B, 5D, 4B, and 4D, respectively, allowing the location of associated functional markers (Lopes et al. 2014). An association mapping study using 1,863 DArT markers was used to identify markers for yield and yield components phenotyped under contrasting moisture regimes in USmid-west (Edae et al. 2014). In the present study we used 18,704 high-density, polymorphic SNPs from the 90K Illumina iSelect SNP array (Wang et al. 2014) to identify molecular markers associated with yield and related traits, with the ultimate aim of facilitating molecular breeding and the strategic combination of traits in spring wheat.

Materials and methods

Plant material

The WAMI population is a genetically diverse collection comprising 287 high-yielding, advanced elite lines of spring wheat. It represents 25 years of research at CIMMYT and was carefully assembled to avoid the confounding effects of phenology and PH in GWAS (Lopes et al. 2012a).

Phenotyping

The WAMI population was evaluated over 4 years (2010–13) under optimal management at the CIMMYT experiment station near Ciudad Obregón, Sonora State, in Northwest Mexico (27.20°N, 109.54°W, 38 masl). This site is a temperate high-radiation environment with adequate irrigation. The trials were timely sown with full irrigation applied through gravity flood-irrigation. In addition, four auxiliary gravity flood-irrigations were also given at regular intervals. Details of date of planting and weather data are shown in Table 1. Phenotypic measurements included grain yield/m^2 (YLD), thousand kernel weight (TKW), grain number/m^2 (GNO) (estimated from YLD and TKW), plant height (PH), days to heading (DTH), days to anthesis (DTA), days to maturity (DTM), chlorophyll index (SPAD) at vegetative stage (SPADvg), SPAD at the grain filling stage (SPADLLg), canopy temperature at vegetative stage (CTvg), canopy temperature at grain-filling (CTLLg), normalized difference vegetation index (NDVI) at vegetative stage (NDVIvg) and at the grain filling stage (NDVILLg), peduncle length (PL), biomass at harvest (BM), and harvest index (HI). For details on measurements and time of measurements please see the manual "Physiological breeding II: a field guide to wheat phenotyping" (Pask et al. 2012).

Genotyping and SNP calling

Seeds of all lines were obtained from the CIMMYT genetic resources program and genomic DNA was extracted from five bulked leaves using a CTAB procedure (Saghai-Maroof et al. 1984) modified as shown in CIMMYT laboratory

protocols (Dreisigacker et al. 2013). DNA was sent for SNP genotyping to the USDA-ARS Small Grain Genotyping Center, Fargo (http://wheat.pw.usda.gov/GenotypingLabs) for use in the Illumina iSelect 90K SNP Assay (Wang et al. 2014), following the manufacturer's protocol. SNP allele clustering and genotype calling was performed with Genome Studio software v2011.1. The default clustering algorithm implemented in genome studio was first used to identify assays that produced three distinct clusters corresponding to the AA, AB, and BB genotypes expected for bi-allelic SNPs. Manual curation was performed for assays that produced compressed SNP allele clusters and could not be discriminated using the default algorithm. The accuracy for SNP clustering was validated visually.

Linkage disequilibrium, population structure, and trait analysis

Linkage disequilibrium (LD) among markers was calculated using the full matrix and sliding window options in TASSEL using 18,704 markers, for each chromosome and for the A, B, and D genomes. For LD calculation, only markers with known position and with a minor allele frequency >5 % were used. Pair-wise LD was measured using the squared allele-frequency correlations r^2, according to Weir (1996). The percentage of marker pairs below and above the critical LD was determined for each chromosome and LD decay was also compared. These results were also compared with Edae et al. (2014) and Lopes et al. (2014).

Population structure of the WAMI population was assessed with 887 SNP markers positioned at least 5 cM apart in the genome selected based on LD decay analysis. Population structure analysis using structure (Pritchard et al. 2000), principal component analysis (PCA) (Pearson 1901), and neighbor joining (NJ) tree analysis (Saitou and Nei 1987) in the Powermarker (Liu and Muse 2005) were additionally carried out and compared with Edae et al. (2014) and Lopes et al. (2014). The structure program was run ten times for each subpopulation (k) value, ranging from 1 to 15, using the admixture model with 20,000 replicates for burn-in and 5,000 replicates during analysis. Kinship was calculated with SPAGeDi 1.3 (Loiselle et al. 1995; Hardy and Vekemans 2002). We also analyzed trait differences among the lines based on 1B/1R translocations that formed major subgroups.

Phenotypic analysis

Analysis of variance and estimates of repeatability were calculated using the proc mixed procedure in SAS 9.1 software (SAS Institute 2000). The experimental design (i.e., alpha lattice) was adjusted considering environments, replicates within environments, incomplete blocks within environments, replications, genotypes, and genotype-by-environment interactions ($G \times E$) as random effects. Adjusted means were calculated for each trait by combining data from four environments and, as relevant for specific traits, using DTH and PH as covariates. Repeatability estimates were calculated in a method similar to that for heritability estimates, using the formula:

$$r^2 = \frac{\sigma^2 G}{\sigma^2 G + \frac{\sigma^2 G \times E}{l} + \frac{\sigma^2 E}{rl}}.$$

where r^2 is the repeatability estimate, $\sigma^2 G$ is the genetic variance, $\sigma^2 G \times E$ is the genotype-by-environmental variance, and $\sigma^2 E$ is the residual variance, r is the number of replications, and l is the number of environments.

Genome-wide association analysis

A dataset including 285 lines was obtained after combining phenotypic (287) and genotypic data (285). Genome-wide scans in TASSEL 4.0 (Bradbury et al. 2007) using 18,704 markers with known positions were conducted using population structure (Q_2–Q_6) as the fixed component and K matrix (kinship matrix) as the random component after model testing. We compared different models in SAS for GWAS to select the best model for detecting marker–trait associations (MTA) following previously recommended procedures for unified mixed model association mapping (Yu et al. 2006; Zhu et al. 2008). Model comparison was done in SAS using the bayesian information criteria (BIC) to select the best model for testing marker trait associations. Simple model, population structure as a cofactor (Q) (Pritchard et al. 2000), K matrix as random term in the mixed model, model involving population structure and familial relatedness ($Q + K$) (Yu et al. 2006), and principal components (PCs) from principal component analysis (PCA) and PCs + K were tested. Results from TASSEL were further verified in SAS applying unified mixed model analysis (Yu et al. 2006). We also tested markers for yield by using DTH and PH as covariates in TASSEL and in estimating LS means for yield. The threshold for defining a marker to be significant was taken at 10^{-04}, considering the number of markers and the deviation of the observed F test statistics from the expected F test distribution (Sukumaran et al. 2012). For traits YLD, TKW, GNO, CTvg, CTLLg, SPADvg, SPADLLg, NDIVvg, and NDVIllg, DTH was used as a covariate in TASSEL due to the phenotypic correlation with phenology traits. GAPIT was also run with the model selection option to check the consistency of the results (Lipka et al. 2012).

Table 2 Descriptive statistics and variance parameters estimated for YLD and yield components studied on 287 lines wheat association mapping initiative (WAMI) population grown in Mexico from 2009–2010 to 2012–2013

Trait	Descriptive statistics			Variance parameters			
	Min	Mean	Max	$\sigma^2 G$	$\sigma^2 (G \times E)$	$\sigma^2 E$	Repeatability
YLD (ton/ha)	4.73	6.79	8.20	0.15	0.11	0.54	0.62
TKW (g)	34.25	42.50	53.40	12.95	1.09	7.08	0.92
GNO	10,890.00	16,139.96	20,827	1,994,636.00	595,889.00	4,005,415.00	0.75
Hed[a] (days)	75.71	82.01	88.82	5.65	1.72	1.99	0.81
Ant (days)	82.58	87.92	95.12	4.20	1.56	5.35	0.75
Mat (days)	125.87	130.35	135.37	1.99	0.63	5.62	0.70
PH (cm)	89.65	103.35	118.55	27.75	2.29	22.83	0.89
SPADLLg	38.29	44.95	50.37	2.62	0.04	4.22	0.55
SPADvg	39.50	43.82	48.43	0.0817	–	4.4302	–
CTvg	21.66	23.10	24.17	0.0002	0.0349	0.6541	–
CTLLg	23.57	24.60	25.56	0.0202	0.0494	0.3226	0.22
NDVIvg	0.30	0.41	0.51	0.0003	0.0009	0.0022	0.23
NDVILLg	0.32	0.44	0.52	0.0007	0.0005	0.0010	0.59
PL[a] (cm)	27.16	35.49	44.11	8.68	1.0224	4.3677	0.84
BM[a]	13.18	17.26	21.36	0.11	0.0123	4.3857	0.09
HI[a]	0.36	0.44	0.49	0.0002	0.0001	0.0009	0.40

[a] Two-year data

$\sigma^2 G$ genotypic variance, $\sigma^2 G \times E$ genotype × environment variance, $\sigma^2 E$ residual variance, *YLD* grain yield, *TKW* thousand kernel weight, *GNO* grain number/m^2, *HED* days to heading, *ANT* days to anthesis, *MAT* days to maturity, *PH* plant height, *SPADLLG* SPAD meter reading at grain filling, *SPADvg* SPAD reading at vegetative state, *CTvg* canopy temperature at vegetative stage, *CTLLg* canopy temperature at grain filling stage *NDVIvg* normalized differential vegetative index at vegetative stage, *NDVILLg* NDVI at grain filling stage, *PL* peduncle length, *BM* biomass, *HI* harvest index

Results

General performance of the 287 lines

Weather conditions, date of planting and emergence, minimum and maximum temperatures, and rainfall were recorded (Table 1). Planting was always in November and environmental conditions were not drastically different across years. Analysis of variance was conducted and minimum, mean, maximum, variance parameters, and repeatability estimates of all the traits studied were estimated (Table 2). Except for SPAD and CT at vegetative stage, all traits were significantly different among genotypes and showed medium-to-high repeatability. The phenotypic ranges for DTH, DTA, and DTM were 13, 13, and 10 days, respectively. Even with a narrow range these traits exhibited higher repeatability estimates. TKW (0.92), had the highest repeatability estimate, followed by PH (0.89), PL (0.84), and DTH (0.81). The lowest repeatability was for BM (0.09). YLD of this population varied from 4.7 to 8.2 ton/ha and TKW from 34.3 to 53.4 g/1,000 kernels. PH showed a range of 28 cm. SPADvg, SPADLLg, CTvg, CTLLg, NDVIvg, and NDVILLg showed low-to-medium repeatability.
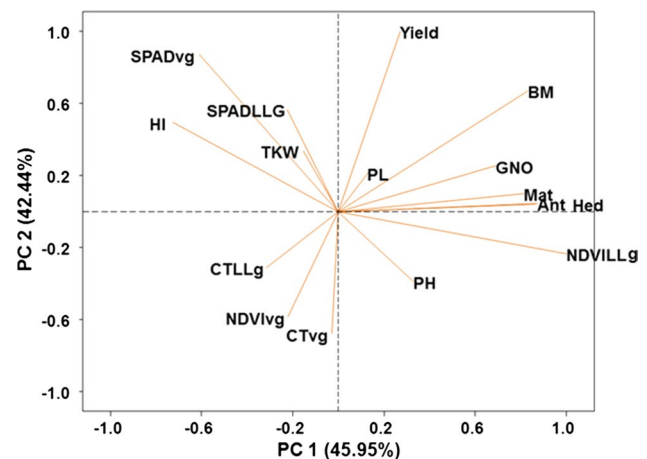

Fig. 1 Bi-plot of the genetic correlation matrix for the traits studied in the wheat association mapping initiative population grown under irrigation at Ciudad Obregón, Mexico (2010–2013)

Bi-plots showing the genetic correlations between different traits were plotted (Fig. 1). YLD was positively correlated with BM, HI, GNO, DTM, NDVILLg, and SPAD. YLD was negatively correlated with CT and PH. TKW and GNO were correlated with DTH and DTA. Pearson

correlation coefficients were calculated for all the trait combinations (Supplementary Table 1). GNO (0.56) and BM (0.53) showed the highest correlation with YLD, followed by HI (0.35). TKW was negatively correlated with GNO (−0.70), but positively correlated with PL (0.48). PL and PH were positively correlated (0.63). NDVILLg was highly correlated with DTM (0.46).

Molecular markers and genetic map

A total of 21,871 (26.8 %) assays showed three distinct clusters corresponding to the AA, AB, and BB genotypes expected for a biallelic SNP. Of the remaining assays with poorer cluster separation, manual clustering was applied. Overall, 36,133 (44.2 %) of the 81,587 functional iSelect bead chip assays visually revealed polymorphism among the WAMI population. We were able to locate 28,614 (35.1 %) of the SNPs on the published genetic map (Wang et al. 2014). The total length of the map was 3,662.67 cM. The number of markers, map length and derived marker density for each chromosome are shown in Supplementary Table 2. Chromosome 4D had the lowest number of markers (186) and chromosome 1B had the highest number of markers (2,456). The map

length was shortest for chromosome 4B (119) and longest for chromosomes 7A and 7D (241). Marker density was lowest for chromosome 6D, with an average spacing of 0.49 cM, and highest for 1B and 6B, with average marker spacing of 0.07 cM. The B genome had the highest number of markers (13,243), followed by the A genome (10,064) and D genome (3,507). Marker density was also highest in the B genome (1 SNP per 0.08 cM), followed by the A genome (1 SNP per 0.12 cM) and the D genome (1 SNP per 0.37 cM).

Linkage disequilibrium, population structure, and trait analysis

We found the highest LD in the D genome. LD decayed below $r^2 = 0.02$ at about 5 cM in the D genome and at about 2 cM in the A and B genomes. A comparison of marker pairs with $r^2 > 0.02$ for all chromosomes indicated that 1A (46 %), 2D (39 %), and 1B (25 %) had the highest percentage of markers in LD. The fewest marker pairs in LD with $r^2 > 0.02$ were on 7D (6 %), 7A (7 %), and 5A (7 %). The D genome had the highest percentage of markers in LD (18 %) followed by the A (16 %) and B genomes (13 %).
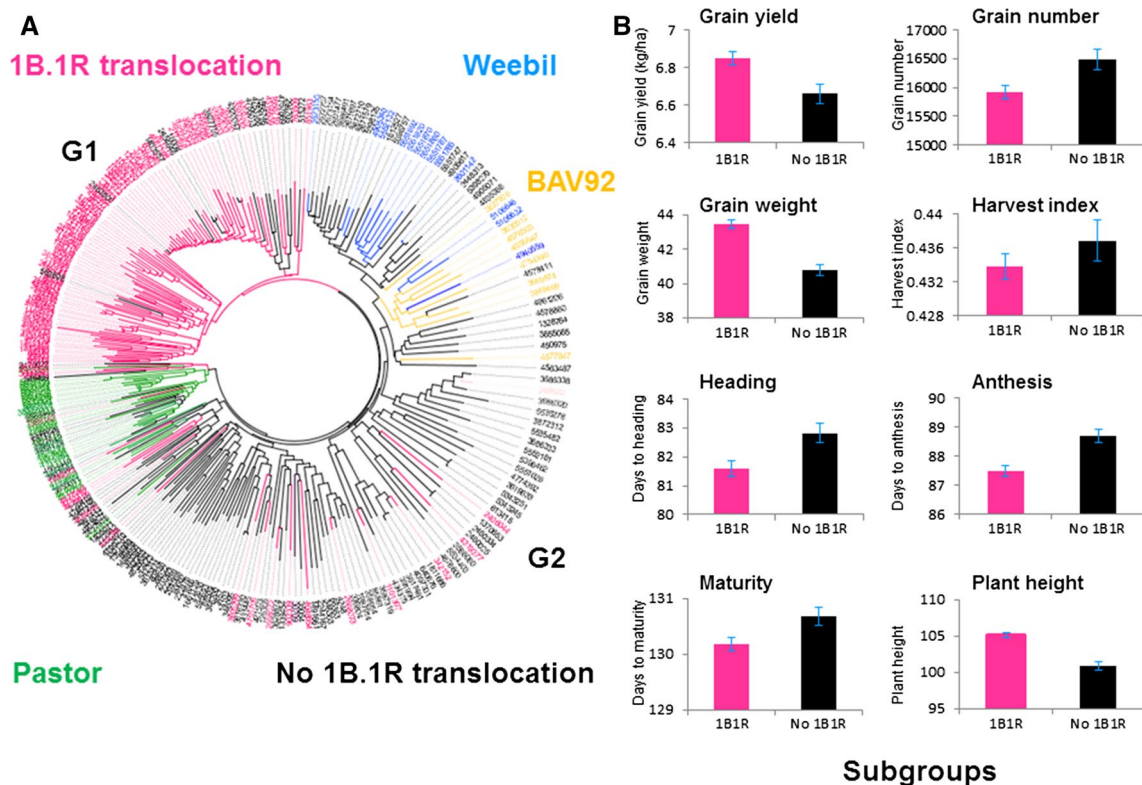

Fig. 2 Trait analysis of the WAMI population based on the 1B/1R translocation and indicating differences between the subpopulations comprising lines with and without the translocation: **a** NJ tree indicating the subpopulations; *lines in red* are with 1B.1R translocation and all others are without 1B.1R translocation **b** differences in the traits between the subpopulations

Population structure analysis using 887 randomly selected SNPs through structure, PCA, and NJ tree analysis indicated that the WAMI population could be divided into two major subpopulations, based the presence or absence of 1B/1R translocations. Clustering according to pedigree included advanced lines derived from diverse, globally important CIMMYT parents and cultivars such as 'Weebil', 'Pastor', and 'BAV92'. Out of 287 lines with genotypic data, 112 lines have 1B/1R translocations in their genomes. The 1B/1R translocation lines showed higher YLD, TKW, and PH, but lower GNO, DTH, DTA, DTM, and HI than lines without the translocation (Fig. 2).

Marker–trait associations (MTAs)

Thirty-one significant marker–trait associations were detected using 18,704 SNPs, after removing alleles with MAF < 5 % from the 26,814 SNP data set. A significance threshold level of 10^{-04} was deemed suitable, considering the deviation of the observed test statistics $[-\log_{10}(p)]$ values from the expected test statistics values in the Q–Q plots (Sukumaran et al. 2012). Testing was done to select the best model for each trait for marker–trait associations (MTAs). The p values using the simple model and the best model for each trait were reported. Significant markers for the traits are shown in Table 3. Manhattan plots of the GWAS results are also shown in Figs. 3 and 4 and in Supplementary Figures 1, 2, and 3.

For YLD, we detected MTAs in chromosomes 3B, 5A, 5B, and 6A. PC2 as the fixed term was the best model for YLD. Markers individually explained 8–11 % of the variation in the trait, and together explained 29 % of the variation in YLD, based on multiple regression analysis. We detected MTAs for TKW on 5A and 6A, using PC2 + K as the best model. The marker *RFL_Contig4632_1512* on 6A was 1 cM away from the marker for YLD on 6A. The marker for grain number on 5A was at 98 cM and close to the marker detected for TKW on 5A at 98 cM. PC2 + K in the mixed model framework was the best model for testing markers for GNO (Fig. 3).

We used PC2 + K model for DTA and DTH, thereby detecting a major region in 5A at 90 cM associated with these traits. The marker explained 7 and 14 % of the variation for DTA and DTH, respectively. For DTM, MTAs were detected on 2B, 3B, 4B, 4D, and 6A that explained 16 % of the variation in the traits. For PH, markers were detected on 1A and 6A that explained 20 % of trait variation, based on multiple regression analysis (Fig. 4).

MTAs were also detected for SPADLLg (3B and 6A), CTLLg (6A, PVE based on multiple regression = 18 %), NDVIvg (4D and 5B), NDVILLg (1B and 6B) that explained 6–7 % of the variation in the traits. MTAs for PL (1B and 3A), BM (3D and 6A, PVE = 12 %), and HI (1B,

1D, and 3B, PVE = 12 %) were also detected (Supplementary Figures 1, 2, and 3).

We detected the largest number of MTAs in chromosome 6A at the 71–85 cM interval, a pleiotropic region for YLD, TKW, GNO, PH, SPADLLg, CTLLg, and BM. Markers detected on 5A were also related to several traits in this population. The markers detected were further verified for allele substitution effect and presented in Supplementary Figures 4 and 5.

Discussion

Trait associations, linkage disequilibrium, population structure, and trait analysis

YLD was positively correlated with TKW, BM, HI, PL, and SPAD readings, as previously reported (Dodig et al. 2012; Lopes et al. 2012c), and negatively correlated with canopy temperature (Jackson et al. 1977; Rashid et al. 1999; Saint Pierre et al. 2010) and NDVI at vegetative stage. Our results on LD using the 18,704 SNP markers were similar to the earlier results from the 9K SNP data (Lopes et al. 2014) and DArT markers (Edae et al. 2014), including finding LD to be greatest in the D genome. Our LD decay rate is about 5 cM for the D genome, which is comparable to that (~6.8 cM) found by Edae et al. (2013). The reason for the high D genome LD may be the introduction of synthetic and new haplotypes in the WAMI population and also due to the evolutionary history of wheat D genome (Edae et al. 2014; Lopes et al. 2014). This indicates that fewer markers are needed for GWAS in the D genome than for the A and B genomes.

We also reached a similar conclusion about population structure using 887 SNPs markers from the 90K SNP data. Structure, PCA, and NJ tree analysis indicated the presence of two subpopulations in this WAMI panel based on the 1B/1R rye translocation (Edae et al. 2014; Lopes et al. 2014). 1B/1R translocation lines had higher YLD, TKW, and early DTH, DTA, and DTM in this WAMI population. Rye-derived wheat cultivars have the short arm of the 1R rye chromosome substituted for the 1B wheat chromosome. This translocation has improved resistance to stem rust, leaf rust, and stripe rust (Dhaliwal et al. 1987; Pena et al. 1990; Wieser et al. 2000).

Earlier studies on the effects of the 1B/1R translocation lines had reported that 1B/1R lines had higher root-and shoot dry weight under drought treatments and an increased root/shoot ratio (Hoffmann 2008). In addition, research has shown increased efficiencies of light conversion, particularly in wheat varieties with the 1B/1R chromosome translocation, hence total crop weights are increased (Sylvester-Bradley et al. 2002). The 1B/1R translocation has been used intensively in wheat breeding but less at CIMMYT in recent

Table 3 Summary of GWAS results

Trait	Marker	Chr.	Position	Alleles	Simple model P value	Best model P value	Marker R^2
YLD	wsnp_Ex_rep_c69664_68618163	3B	71	25 (C)/254 (T)	7.48E−11	6.49E−09	0.11
	Kukri_c2700_1137	6A	85	94 (C)/187 (T)	2.41E−06	1.74E−06	0.07
	Kukri_c14877_303	6A	79	86 (C)/196 (T)	2.41E−08	9.48E−08	0.09
	BS00064336_51	5A	115	23 (A)/261 (G)	4.45E−08	1.79E−07	0.09
	Excalibur_c13408_280	5B	129	94 (T)/178 (C)	2.02E−10	8.18E−07	0.08
TKW	BobWhite_c14689_172	5A	98	134 (T)/139 (C)	1.34E−11	1.56E−06	0.09
	RFL_Contig4632_1512	6A	78	44 (A)/232 (G)	3.59E−09	6.67E−06	0.08
GNO	Ku_c21235_676	5A	98	126 (T)/137 (C)	1.22E−05	5.23E−06	0.08
Heading	Kukri_c9411_676	5A	90	53 (C)/226 (T)	5.12E−19	3.68E−09	0.14
Anthesis	Kukri_c9411_676	5A	90	25 (C)/254 (T)	3.46E−06	1.97E−05	0.07
Maturity	Kukri_rep_c112440_422	3B	82	108 (G)/162 (A)	2.82E−05	2.83E−05	0.06
	BobWhite_c30930_192	6A	71	119 (C)/166 (T)	4.98E−05	4.98E−05	0.05
	wsnp_Ex_rep_c107564_91144523	4D	70	58 (T)/219 (C)	6.31E−05	6.31E−05	0.05
	Kukri_c10869_489	2B	145	112 (C)/169 (T)	7.97E−05	7.97E−05	0.05
	BS00034371_51	4B	76	53 (C)/189 (T)	7.29E−05	7.29E−05	0.06
PH	wsnp_Ex_c1427_2736441	1A	70	42 (A)/227 (G)	6.45E−11	1.58E−06	0.07
	IAAV7384	6A	80	69 (T)/191 (G)	9.95E−03	6.27E−05	0.06
SPADLLg	Kukri_c48750_1372	3B	61	50 (C)/228 (T)	5.11E−03	2.61E−05	0.06
	wsnp_Ex_c11348_18326787	6A	85	39 (A)/240 (C)	8.20E−04	3.11E−05	0.06
CTLLg	RAC875_rep_c114561_587	6A	74	119 (G)/156 (A)	3.77E−07	2.84E−05	0.07
NDVIvg	wsnp_CAP7_c2086_1018815	5B	43	42 (G)/231 (A)	1.57E−05	2.56E−05	0.06
	Tdurum_contig60051_169	4D	128	89 (C)/183 (T)	2.28E−06	7.13E−05	0.05
NDVILLg	GENE-4086_876	6B	73	95 (G)/181 (A)	5.87E−05	1.40E−05	0.07
	BS00011973_51	1B	76	90 (G)/191 (T)	1.08E−04	2.70E−05	0.06
PL	Tdurum_contig70304_781	1B	70	84 (G)/184 (A)	1.41E−07	1.46E−04	0.05
	BS00039498_51	3A	122	24 (A)/254 (G)	1.70E−04	2.21E−05	0.06
BM	D_F5MV3MU01DWT7N_151	3D	148	6 (T)/278 (C)	1.12E−06	1.12E−06	0.08
	BS00072903_51	6A	85	92 (A)/189 (G)	3.47E−05	3.48E−05	0.05
HI	Kukri_c67675_174	1D	167	111 (T)/157 (C)	8.78E−08	6.58E−06	0.07
	BobWhite_c20073_443	1B	148	109 (G)/160 (A)	2.05E−06	3.53E−05	0.06
	BS00079989_51	3B	14	33 (C)/245 (T)	5.92E−04	6.88E−05	0.05

Chr chromosome, *YLD* grain yield, *TKW* thousand kernel weight, *GNO* grain number/m², *HED* days to heading, *ANT* days to anthesis, *MAT* days to maturity, *PH* plant height, *SPADLLG* SPAD meter reading at grain filling, *SPADvg* SPAD reading at vegetative state, *CTvg* canopy temperature at vegetative stage, *CTLLg* canopy temperature at grain filling stage, *NDVIvg* normalized differential vegetative index at vegetative stage, *NDVILLg* NDVI at grain filling stage, *PL* peduncle length, *BM* biomass, *HI* harvest index

times, due to its negative effect on quality-sticky dough (Malik et al. 2013). The IB/IR translocation donated by a seri parent in the seri/babax recombinant inbred population was also associated with reduced yield under drought and high temperatures (Pinto et al. 2010; Lopes et al. 2014). Screening for the 1B/1R translocation in other germplasm might help in breeding for early flowering, high-yielding lines, but it depends considerably on the genetic background and traits associated with bread-making quality.

Most WAMI genotypes have been crossed extensively to produce elite lines that were used in CIMMYT breeding programs (Lopes et al. 2012a) and resulting in a population with lines related to each other. An earlier paper had reported seven WAMI subpopulations using a small number of DArT markers (Edae et al. 2014). Lopes et al. (2014) used SNP markers to assess WAMI population structure and concluded similar to our study. Principal component analysis using 9,000 GBS markers gave consistent results supported by 1B/1R translocation information (unpublished results).

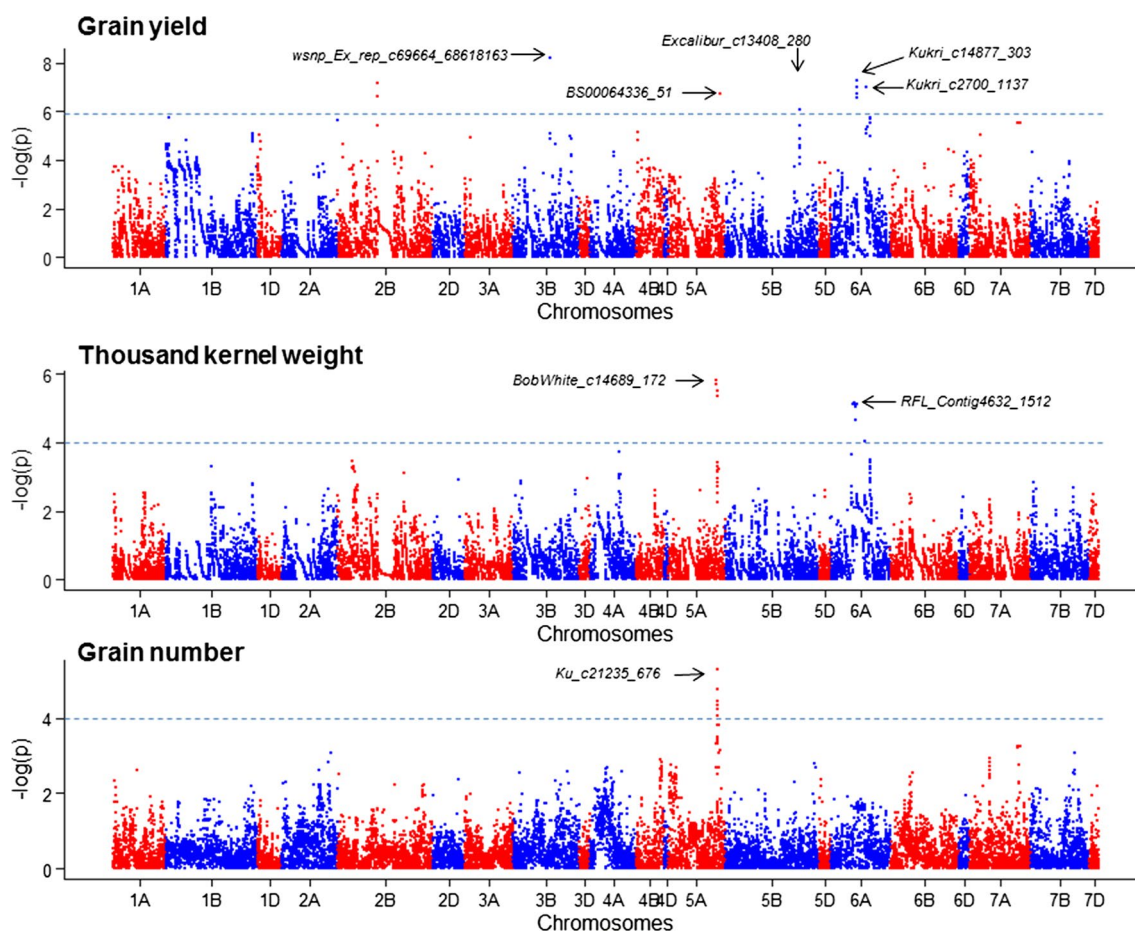In GWAS there are two ways to account for confounding population structure: (1) statistically accounting for

Fig. 3 GWAS results using 18,704 SNPs markers in WAMI for yield traits based on least square means of the combined data from 2010 to 2013. The *blue horizontal line* indicates threshold of significance

population structure effects and (2) carefully selecting the association mapping panel to reduce the range of phenology. In our study, this population was carefully selected to avoid the confounding effects and at the same time, unified mixed model approach (Yu et al. 2006) with model testing for each trait was followed to statistically reduce the confounding effects in association mapping. Blindly using the $Q + K$ model in GWAS has resulted in over-correcting for population structure and, thus, false positives (Yu et al. 2006). DTA and PH were also used as covariates in marker–trait detection in TASSEL, when appropriate and also in estimation of LS means. We did not detect any marker for YLD and yield components confounded with DTH. However, markers for YLD and PH are present in the 6A pleiotropic region.

Marker–trait associations

For YLD, MTA were observed in chromosomes 3B, 5A, 5B, and 6A using DTH and PH as covariates separately and in combination. The marker in chromosome 6A is at

a pleiotropic region affecting YLD, TKW, SPADLLg, PH, and CTLLg. Earlier studies have reported a QTL for YLD in short arm of chromosome 6A in winter wheat (Snape et al. 2007) and spring wheat (Lopes et al. 2014).

In addition to the markers on 6A, a pleiotropic locus for YLD, TKW, and GNO was identified in chromosome 5A at 98 cM. We identified this locus significantly associated with GNO and TKW with and without DTH as cofactor. The marker on 5A was close to the DTH marker detected in 5A at 90 cM, but the significant marker for each trait was different. GNO was correlated with DTH, DTA, and DTM so we used DTH as a covariate to detect markers for GNO and TKW.

Another region significantly associated with YLD in 5B is a multi-trait region significant for yield and yield components in spring wheat (Edae et al. 2014). Our marker was 20 cM away from the marker reported by this previous study. Pinto et al. (2010) also reported a robust QTL for YLD in chromosomes 3B and 5B. Lopes et al. (2014) identified SNP markers in chromosomes 3B and 5A for YLD that were most consistent across heat and drought environments.
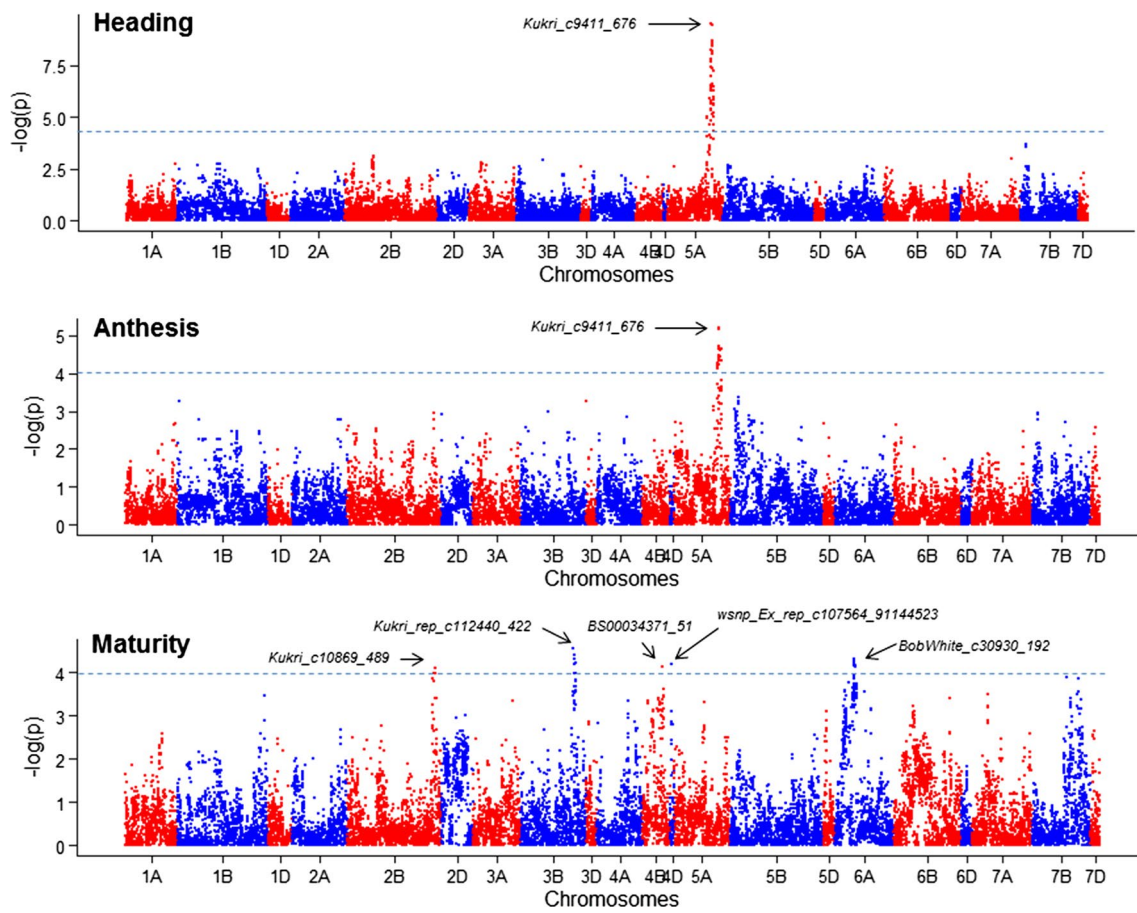
Fig. 4 GWAS results using 18,704 SNPs markers in WAMI for days to heading, anthesis, and maturity (2010–2013)

We identified MTA for DTH in chromosome 5A. The locus associated with DTA was also associated with DTH indicating pleiotropic effect of the gene. An earlier blind association analysis of this population has shown the position of Vrn gene in the 5A chromosome (Lopes et al. 2014). Candidate gene association mapping using the functional gene Vrn-1A indicated that it is associated with DTH and DTA. There were several SNPs significant for DTH and DTA at 90 cM in chromosome 5A indicating that the same region determines DTA and DTH.

Several MTAs were detected for grain maturity in chromosomes 2B, 3B, 4B, 4D, and 6A. Earlier studies have reported QTLs for maturity in chromosomes 2B and 4D in a biparental population (Pinto et al. 2010). Markers for PH were detected in chromosomes 1A and 6A in this study. The locus on 6A was associated with YLD and was consistently detected under drought, heat, and irrigated conditions (Edae et al. 2014; Lopes et al. 2014). As was found in the present study, Edae et al. (2014) did not report any *Rht* gene related to PH in this WAMI population. From the genetic and phenotypic correlation analysis PH was not correlated with YLD, indicating that the loci on 6A have pleiotropic effects on several traits.

MTAs for CTLLg were detected in chromosomes 2D, 5A, and 6A and the MTA on 6A was associated with YLD. Mason et al. (2013) detected QTL for canopy temperature depression in chromosomes 3BL and 5DL that were independent of phenology QTL. Pinto et al. (2010) detected QTLs for canopy temperature in chromosomes 1A, 1B, 1D, 2B, 3B, 4A, 5A, 5B, and 7A. Use of these MTAs to select for lower canopy temperature might help in higher yields (Reynolds et al. 1998; Cossani and Reynolds 2012). The SPADLLg marker in chromosome 3B was 10 cM close to the yield loci detected in our study. The other locus on 6A at 85 cM was pleotropic for YLD and SPADLLg. Two loci in total were detected for NVDI in chromosomes 1B, 4D, 5B, and 6B. Pinto et al. (2010) identified markers for NDVI in chromosomes 1B, 4D, 5B, and 6B, the one on 5B being a pleotropic loci. For peduncle length loci were detected in chromosomes 1B and 3A, and PH was used as a covariate to reduce the confounding effect of PH on PL. Earlier studies have shown co-localization of PL QTLs with PH (Heidari et al. 2012). BM markers were detected on 3D and 6A. Earlier studies have shown the pleiotropic effect of QTLs for BM and YLD (Mason et al. 2013). Three loci

in chromosomes 1B, 1D, and 3B were detected for HI in this study. We did not find earlier studies detecting the similar regions in spring wheat but studies have detected yield QTLs in chromosomes 1B, 1D, and 3B, and that HI was correlated with YLD (Pinto et al. 2010). In an earlier GWAS with DArT markers in this population, markers for HI were detected in chromosomes 5AL and 5B under drought conditions (Edae et al. 2014).

Conclusions

Association mapping is a powerful tool to identify molecular markers for physiological and agronomic traits in wheat. Through GWAS we identified pleiotropic loci in spring wheat associated with YLD, TKW, PH, and physiological measurements. The 1B/1R rye translocation and line pedigrees significantly determined the structure of the WAMI population. Some traits varied in values among the two subpopulations (that is, the one containing lines with the 1B/1R translocation and the one containing lines without). Thirty-one significant loci were detected for YLD and related traits in this population. The loci on 5A and 6A will be further validated for use in breeding programs.

Author contribution statement M.R. designed the research; M.S.L., P.C., S.D., S.S. conducted the research; S.S., S.D. wrote the manuscript.

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

References

Bradbury PJ, Zhang Z, Kroon DE et al (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23:2633–2635. doi:10.1093/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:10.1534/genetics.105.044586

Chao S, Dubcovsky J, Dvorak J et al (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

Cossani CM, Reynolds MP (2012) Physiological traits for improving heat tolerance in wheat. Plant Physiol 160:1710–1718. doi:10.1104/pp.112.207753

Curtis T, Halford NG (2014) Food security: the challenge of increasing wheat yield and the importance of not compromising food safety. Ann Appl Biol. doi:10.1111/aab.12108

Dhaliwal AS, Mares DJ, Marshall DR (1987) Effect of 1B-1R chromosome translocation on milling and quality characteristics of bread wheats. Cereal Chem 64:72–76

Dodig D, Zoric M, Kobiljski B et al (2012) Genetic and association mapping study of wheat agronomic traits under contrasting water regimes. Int J Mol Sci 13:6167–6188. doi:10.3390/ijms13056167

Dreisigacker S, Tiwari R, Sheoran S (2013) Laboratory manual: ICAR-CIMMYT molecular breeding course in wheat. ICAR/BMZ/CIMMYT, Haryana, India, p 36

Edae EA, Byrne PF, Manmathan H, Haley SD, Moragues M, Lopes MS, Reynolds MP (2013) Association mapping and nucleotide sequence variation in five drought tolerance candidate genes in spring wheat. Plant Genome 6(2). doi:10.3835/plantgenome2013.04.0010

Edae EA, Byrne PF, Haley SD, Lopes MS, Reynolds MP (2014) Genome-wide association mapping of yield and yield components of spring wheat under contrasting moisture regimes. Theor Appl Genet 127(4):791–807. doi:10.1007/s00122-013-2257-8

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620. doi:10.1046/j.1471-8286.2002.00305.x

Heidari B, Saeidi G, SayedTabatabaei BE, Suenaga K (2012) QTLs involved in PH, peduncle length and heading date of wheat (*Triticum aestivum* L.). J Agric Sci Technol 14(5):1090–1104

Hoffmann B (2008) Alteration of drought tolerance of winter wheat caused by translocation of rye chromosome segment 1RS Cereal Research Communications. Cereal Res Commun 36(2):269–278. doi:10.1556/CRC.36.2008.2.7

Jackson RD, Reginato RJ, Idso SB (1977) Wheat canopy temperature: a practical tool for evaluating water requirements. Water Resour Res 13:651. doi:10.1029/WR013i003p00651

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28(18):2397–2399

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. doi:10.1093/bioinformatics/bti282

Loiselle BA, Sork VL, Nason JD, Graham C (1995) Spatial genetic structure of a tropical understory shrub, *Psychotria officinalis* (Rubiaceae). Am J Bot 82:1420–1425. doi:10.2307/2445869

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

Lopes MS, Reynolds MP, Manes Y et al (2012b) Genetic yield gains and changes in associated traits of CIMMYT spring bread wheat in a "historic" set representing 30 years of breeding. Crop Sci 52:1123. doi:10.2135/cropsci2011.09.0467

Lopes MS, Reynolds MP, McIntyre CL et al (2012c) 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. doi:10.1007/s00122-012-2030-4

Lopes MS, Dreisigacker S, Peña RJ, Sukumaran S, Reynolds MP (2014) Genetic characterization of the Wheat Association Mapping Initiative (WAMI) panel for dissection of complex traits in spring wheat (submitted)

Malik R, Tiwari R, Arora A, Kumar P, Sheoran S, Sharma P, Singh R, Sharma I (2013) Genotypic characterization of elite Indian wheat genotypes using molecular markers and their pedigree analysis. Aust J Crop Sci 7(5):561

Mason RE, Hays DB, Mondal S et al (2013) QTL for yield, yield components and canopy temperature depression in wheat under late sown field conditions. Euphytica 194:243–259. doi:10.1007/s10681-013-0951-x

Pask AJD, Pietragalla J, Mullan, Reynolds MP (eds) (2012) Physiological breeding II: a field guide to wheat phenotyping. International Wheat and Maize Improvement Centre (CIMMYT), DF, Mexico

Pearson K (1901) LIII. On lines and planes of closest fit to systems of points in space. Philos Mag Ser 6(2):559–572. doi:10.1080/14786440109462720

Pena RJ, Amaya A, Rajaram S, Mujeeb-Kazi A (1990) Variation in quality characteristics associated with some spring 1B/1R translocation wheats. J Cereal Sci 12:105–112. doi:10.1016/S0733-5210(09)80092-1

Pinto RS, Reynolds MP, Mathews KL et al (2010) Heat and drought adaptive QTL in a wheat population designed to minimize confounding agronomic effects. Theor Appl Genet 121:1001–1021. doi:10.1007/s00122-010-1351-4

Poland JA, Bradbury PJ, Buckler ES, Nelson RJ (2011) Genome-wide nested association mapping of quantitative resistance to northern leaf blight in maize. Proc Natl Acad Sci USA 108:6893–6898. doi:10.1073/pnas.1010894108

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959. doi:10.1111/j.1471-8286.2007.01758.x

Rashid A, Stark JC, Tanveer A, Mustafa T (1999) Use of canopy temperature measurements as a screening tool for drought tolerance in spring wheat. J Agron Crop Sci 182:7. doi:10.1046/j.1439-037x.1999.00335.x

Reynolds MP, Singh RP, Ibrahim A et al (1998) Evaluating physiological traits to complement empirical selection for wheat. Euphytica 100:85–94. doi:10.1023/A:1018355906553

Risch N, Merikangas K (1996) The Future of Genetic Studies of Complex Human Diseases. Science 80(273):1516–1517. doi:10.1126/science.273.5281.1516

Saghai-Maroof MA, Soliman KM, Jorgensen RA, Allard RW (1984) Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. Proc Natl Acad Sci 81(24):8014–8018

Saint Pierre C, Crossa J, Manes Y, Reynolds MP (2010) Gene action of canopy temperature in bread wheat under diverse environments. Theor Appl Genet 120:1107–1117. doi:10.1007/s00122-009-1238-4

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425. doi:0737-4038/87/0404-0007$02.00

SAS Institute (2000) SAS 9.1.3 help and documentation. SAS Institute Inc, Cary

Sharma RC, Crossa J, Velu G et al (2012) Genetic gains for YLD in CIMMYT spring bread wheat across international environments. Crop Sci 52:1522. doi:10.2135/cropsci2011.12.0634

Snape JW, Foulkes MJ, Simmonds J et al (2007) Dissecting gene × environmental effects on wheat yields via QTL and physiological analysis. Euphytica 154:401–408. doi:10.1007/s10681-006-9208-2

Sukumaran S, Yu J (2014) Association mapping of genetic resources: achievements and future perspectives. In: Genomics of plant genetic resources, Springer, Netherlands, pp 207–235. doi:10.1007/978-94-007-7572-5_9

Sukumaran S, Xiang W, Bean SR et al (2012) Association mapping for grain quality in a diverse sorghum collection. Plant Genome J 5:126. doi:10.3835/plantgenome2012.07.0016

Sylvester-Bradley R, Lunn G, Foulkes J, Shearman V, Spink J, Ingram J (2002) Management strategies for high yields of cereals and oilseed rape. In: HGCA R and D conference-agronomic intelligence: the basis for profitable production, pp 8–1

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang E, Maccaferri M, Salvi S, Milner S, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, IWGSC, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo M, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90,000 SNP array. Plant Biotechnol J (in press)

Weir BS (1996) Genetic data analysis II: methods for discrete population genetic data. Sinauer Assoc Sunderl MA 376. doi:10.1111/j.1365-2052.2009.01970.x

Wieser H, Kieffer R, Lelley T (2000) The influence of 1B/1R chromosome translocation on gluten protein composition and technological properties of bread wheat. J Sci Food Agric 80:1640–1647. doi:10.1002/1097-0010(20000901)80:11<1640:AID-JSFA688>3.0.CO;2-4

Würschum T, Langer SM, Longin CFH et al (2013) Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. Theor Appl Genet 126:1477–1486. doi:10.1007/s00122-013-2065-1

Yu J, Pressoir G, Briggs WH et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208. doi:10.1038/ng1702

Yu L-X, Lorenz A, Rutkoski J et al (2011) Association mapping and gene-gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theor Appl Genet 123:1257–1268. doi:10.1007/s00122-011-1664-y

Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. Plant Genome J 1:5. doi:10.3835/plantgenome2008.02.0089