

RESEARCH ARTICLE

# Association of Agronomic Traits with SNP Markers in Durum Wheat (*Triticum turgidum* L. *durum* (Desf.))

Xin Hu<sup>1</sup>, Jing Ren<sup>2</sup>, Xifeng Ren<sup>1</sup>, Sisi Huang<sup>1</sup>, Salih A. I. Sabiel<sup>1</sup>, Mingcheng Luo<sup>3</sup>, Eviatar Nevo<sup>4</sup>, Chunjie Fu<sup>5</sup>, Junhua Peng<sup>5\*</sup>, Dongfa Sun<sup>1,6\*</sup>

**1** College of Plant Science and Technology, Huazhong Agricultural University, Wuhan Hubei, 430070, China, **2** Shandong Provincial Key Laboratory of Functional Macromolecular Biophysics, Institute of Biophysics, Dezhou University, Dezhou, Shandong, 253023, China, **3** Department of Plant Sciences, University of California Davis, Davis, CA, 95616, United States of America, **4** Institute of Evolution, University of Haifa, Mount Carmel, Haifa, 31905, Israel, **5** Science and Technology Center, China National Seed Group Co., Ltd, Wuhan, Hubei, 430206, China, **6** Hubei Collaborative Innovation Center for Grain Industry, Jingzhou, Hubei, 434025, China

\* [jpeng@lamar.colostate.edu](mailto:jpeng@lamar.colostate.edu) (JHP); [sundongfa1@mail.hzau.edu.cn](mailto:sundongfa1@mail.hzau.edu.cn) (DFS)



**OPEN ACCESS**

**Citation:** Hu X, Ren J, Ren X, Huang S, Sabiel SAI, Luo M, et al. (2015) Association of Agronomic Traits with SNP Markers in Durum Wheat (*Triticum turgidum* L. *durum* (Desf.)). PLoS ONE 10(6): e0130854. doi:10.1371/journal.pone.0130854

**Editor:** Wujun Ma, Murdoch University, AUSTRALIA

**Received:** March 18, 2015

**Accepted:** May 25, 2015

**Published:** June 25, 2015

**Copyright:** © 2015 Hu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files. SNP markers were discovered in a panel of 32 lines of tetraploid and hexaploid wheat, and the marker data information is available at a USDA public database, the Wheat SNP Database (<http://probes.pw.usda.gov:8080/snpworld/Search>).

**Funding:** This study was supported by "863" China High-Tech Research Program (2011AA10A106) (DS), The China National Science Foundation (31030055 and 30870233) (JP). The China National Seed Group Co., Ltd. provided support in the form of salaries for authors [CF and JP], but did not have any additional role in the study design, data collection and analysis,

## Abstract

Association mapping is a powerful approach to detect associations between traits of interest and genetic markers based on linkage disequilibrium (LD) in molecular plant breeding. In this study, 150 accessions of worldwide originated durum wheat germplasm (*Triticum turgidum* spp. *durum*) were genotyped using 1,366 SNP markers. The extent of LD on each chromosome was evaluated. Association of single nucleotide polymorphisms (SNP) markers with ten agronomic traits measured in four consecutive years was analyzed under a mix linear model (MLM). Two hundred and one significant association pairs were detected in the four years. Several markers were associated with one trait, and also some markers were associated with multiple traits. Some of the associated markers were in agreement with previous quantitative trait loci (QTL) analyses. The function and homology analyses of the corresponding ESTs of some SNP markers could explain many of the associations for plant height, length of main spike, number of spikelets on main spike, grain number per plant, and 1000-grain weight, etc. The SNP associations for the observed traits are generally clustered in specific chromosome regions of the wheat genome, mainly in 2A, 5A, 6A, 7A, 1B, and 6B chromosomes. This study demonstrates that association mapping can complement and enhance previous QTL analyses and provide additional information for marker-assisted selection.

## Introduction

Durum wheat (*Triticum durum* Desf.) is a tetraploid species consisting of A and B genomes (AABB). It was resulted from domestication of wild emmer wheat (*T. dicoccoides*) derived from a spontaneous cross between *T. urartu* (AA genome, 2n = 14) and an ancient relative of

decision to publish, or preparation of the manuscript. The specific roles of the author are articulated in the 'author contributions' section.

**Competing Interests:** CF and JP are employees in the China National Seed Group Co., Ltd. There are no patents, products in development or marketed products to declare. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

*Aegilops speltoides* (donor of the BB genome) [1]. As the main source of semolina for the production of pasta, bagel, couscous and other Mediterranean local end-products [2], durum wheat is cultivated on about 17 million hectares worldwide. The durum wheat is mainly grown in Europe, Canada, Syria, USA, Algeria and Morocco, particularly in the Mediterranean, while minor grown in Russia, Turkey, Tunisia, Mexico and India [3]. It plays an important role in food production of these regions (<http://www.pasta-unafpa.org/ingstatistics5.htm>).

The amount of genetic variation in germplasm and genetic relationships between genotypes are very valuable information for effective conservation and utilization of genetic resources [4]. Genetic diversity is the foundation for survival, adaptation and evolution in time and space [5]. More knowledge on the genetic variation and the genetic determinants of diversity is useful for discovering new genes [6–9]. Its preservation and wise application in nature are the central aspect of biological conservation and genetic improvement. Assessment of genetic diversity in durum germplasm will provide useful information for breeding programs. Genetic diversity in germplasm can be characterized by different markers: like morphology, pedigree and molecular markers. Currently, application of molecular markers is the most effective and feasible method for characterizing diversity in wild and cultivated germplasm [7, 8].

Genetic diversity analysis of wild and cultivated wheat is generally based on low-to medium-throughput marker platforms such as restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP) and simple sequence repeat (SSR) [10–14]. These molecular markers have been shown to be useful for studying genetic diversity and structure, and differentiating durum wheat cultivars to some extent. However, these markers, especially RFLP, RAPD and AFLP, have not been used extensively in breeding programs because they are not more efficient for application in marker-assisted-selection (MAS) [15].

Single nucleotide polymorphism (SNP) can be converted into genetic markers amenable to high-throughput assays [16, 17]. As the continuous discovery of SNP and the further development of SNP-genotyping platforms, SNP markers gain increasing more and more attentions [18–21]. Genome-wide maps consisting of large number of SNP markers have been reported in *Arabidopsis* [22], rice [23], soybean [24] and barley [25]. However, large-scale SNP detection is restricted by both the polyploidy nature and the high sequence similarity among the three homoeologous genomes in wheat [26, 27]. A relatively small number of SNP markers are now available in wheat due to the genome complexity [18, 28].

Currently, SNP marker is the major type of molecular markers used in evaluating genetic diversity, population structure, familial kinship and associations in multiple organisms. The availability of wheat SNP markers allows trait-marker association analysis with a high efficiency in durum wheat. Association analysis, or association mapping (AM), is a method based on linkage disequilibrium (LD) that is used to detect the relationship between phenotypic variation and genetic polymorphisms [29, 30]. Originally developed for human genetics [31, 32], AM has recently appeared as an alternative approach to mapping QTLs and genes in many crops, due to the development of cheaper, faster and higher density molecular markers [33]. In comparison with genetic linkage analysis, AM has three obvious advantages that include shorter research time, much higher mapping resolution and a greater number of alleles [34]. Relative to other experimental designs that require sampling within families, AM offers the important advantage that allows sampling unrelated individuals in the population for studying genetics of complex traits [29, 35]. AM has been applied to many crop species, such as maize, soybean, rice, barley, wheat etc. [9, 36–38]. Therefore, AM provides a powerful tool for investigating genetics of quantitative traits in plant species [34, 39, 40].

Genetic diversity and association analysis in wheat germplasm have been studied using several types of molecular markers including SNP [9]. Recently, there are more reports on

diversity pattern and population structure in durum wheat germplasm [41–43]. However, there are few reports published on trait-marker associations in durum wheat. Therefore, the major objective of this study is to reveal associations between quantitative traits and the SNP markers in durum wheat.

## Materials and Methods

### Plant materials and field trials

One hundred and fifty durum wheat accessions of worldwide origin were investigated in this study. The collection of the durum wheat germplasm was classified into seven groups based on their geographic origins. Of the accessions, 24 originated from West Asia (WA), 25 from East Asia (EA), 33 from North America (NA), 33 from different parts of Europe (EU), 12 from South America (SA), 16 from North Africa (AF), and 7 from Australia (AU). The name, place of origin and identifier number for each accession is listed in [S1 Table](#).

In order to obtain reliable phenotypic data, field trials of all the accessions with replications were conducted in four consecutive years. The field trials got the approval of Huazhong Agricultural University, and were performed on the experimental farm of Huazhong Agricultural University, Wuhan, China. The land accessed is not privately owned nor protected, which is belong to Huazhong Agricultural University. All of the materials used in this study were acquired by Dr. Junhua Peng from USDA (United States Department of Agriculture), and no any protected species were sampled in the field trials. The trials with three replications were planted around the end of October in 2009, 2010, 2011 and 2012, respectively, in two rows with 1 m in length and 20 cm between rows, 6 plants in each row. Because some of the accessions were very tall and easy-lodging, we installed frames made of bamboo sticks in each plot before heading to prevent lodging, or reduce lodging impact on the traits.

### Phenotyping of the key traits

**Measurement of key traits.** After full maturity, we randomly harvested four individual plants from each plot. The following 10 traits were measured. The mean value of a trait in each replication was calculated.

PH: plant height (cm),

ES: number of effective spikes,

LMS: length of main spike (cm),

SMS: number of spikelets on main spike,

RLMS: rachis internode length of main spike (cm),

NSPP: number of spikelets per plant,

LFPMS: panicle neck length of main spike (cm),

GNP: grain number per plant,

GWP: grain weight per plant (g),

KGW: 1000-grain weight (g).

**Variation analysis.** The mean phenotypic values of the 10 quantitative traits were subjected to statistical analysis. Frequency distribution of the traits was analyzed, and

Kolmogorov–Smirnov test was performed to test for normal distribution. Data transformation is performed for the traits that did not fit the normal distribution. Calculations of the descriptive statistics, analysis of variance (ANOVA) and broad-sense heritability ( $H^2$ ), and correlation analysis were performed using SPSS programs (IBM SPSS Statistics, Chicago, IL, USA).

## DNA extraction, SNP genotyping and marker data analyses

Before the elongation stage of wheat plants, approximately 1.0 g of young leaf tissue was collected from each of the accessions. The tissue was placed in a 1.5 ml Eppendorf tube, immediately frozen in liquid N, and stored in a  $-80^{\circ}\text{C}$  freezer [43]. The cetyltrimethyl ammonium bromide (CTAB) method was used to extract the total genomic DNA [44].

The DNA samples were shipped to University of California at Davis, USA for genotyping. A set of 1,536 genome-specific SNP markers were applied to genotype the germplasm. These SNP markers were discovered in a panel of 32 lines of tetraploid and hexaploid wheat (<http://avena.pw.usda.gov/SNP/internal/protocol/id.htm>), and downloaded from the Wheat SNP Database (<http://probes.pw.usda.gov:8080/snpworld/Search>). The SNP-genotyping was performed using the Illumina Bead Array platform and Golden Gate Assay (Illumina, San Diego, CA) at the UC Davis Genome Center (<http://www.genomecenter.ucdavis.edu/dna-technologies>). The SNP markers were treated as co-dominant markers. The details of genotyping and genetic analyses were described in Ren et al. [43].

## Linkage disequilibrium

It is essential for association mapping to examine the degree of LD in the genome and chromosome [45,46]. The fraction of locus pairs indicating significant LD increases with decreasing significance level. A high significance level of  $p < 0.001$  was chosen for comparative purposes. If all pairs of adjacent loci within a chromosomal region were in significant LD, this region was treated as a LD block [47]. LD between markers was measured using  $R^2$ , square of correlation between the markers [48]. The values of  $R^2$  and  $P$  were calculated using the software TASSEL 3.0.124 (<http://www.maizegenetics.net/>).

## Association analysis

Association mapping analysis between SNP markers and the 10 quantitative traits (PH, ES, LMS, SMS, RLMS, LFPMS, NSPP, GNP, GWP, and KGW) was performed based on the general linear model (GLM) and the mixed linear model (MLM) using software TASSEL 3.0.124 (<http://www.maizegenetics.net/tassel>). The population structure was estimated using STRUCTURE 2.3.4 software [49] as in Ren et al. [43]. The pair-wise kinship coefficients were estimated according to the method of Lynch and Ritland [50], performed in the program SPAGeDi [51] (<http://ebe.ulb.ac.be/ebe/SPAGeDi.html>). The number of permutation runs was set as 10,000 to obtain the permutation-based significance in GLM analysis. MLM was fitted for each marker and phenotype, accounting for Q-Matrix of the population structure as a covariate and pair-wise kinship coefficients (K matrix) as random effects [34]. Significance of associations between marker loci and traits was tested at a corresponding level of the experiment-wise P-value. Significance of associations between loci and traits was described as P-value and the QTL effects were evaluated by marker- $R^2$  [52].

## Results

### SNP markers and population structure

Multiplexed 1,536 Illumina Golden Gate SNP assay involving in 150 durum wheat accessions generated 230,400 data points. Out of the examined SNPs, 1,366 (89%) were successfully amplified, and other 10% were missing. The detailed analyses on the SNP markers were reported in Ren et al. [43]. The SNP loci were well distributed across the seven homoeologous chromosome groups. The total marker number ranged from 161 in group 5 to 236 in group 7 chromosomes. The number of polymorphic markers ranged from 108 in group 5 to 161 loci in group 6 chromosomes [43].

The structure analysis was performed in Ren et al. [43], and the result suggested that the observed durum wheat germplasm can be divided into two genetically distinct groups (Group I and Group II). The cluster analysis showed that the group II can be further divided into four subgroups, IIa, IIb, IIc, and IId. The dendrogram of 150 durum wheat landraces based on the shared-allele genetic distance calculated from 1,366 SNP markers was showed in Ren et al. [43].

### Linkage disequilibrium among intra-chromosome SNP loci

A total of 1,338 SNP markers with a mean marker density of 95–96 markers per chromosome, ranging from 66 (3B) to 130 (7A) for all the 14 chromosomes, were used to calculate the extent of LD. The pattern of LD was measured using  $R^2$  of allele pairs between 2 loci according to Weir and Cockerham [53] on both chromosome and genome levels (Tables 1 and 2).

**Table 1. SNP locus pairs on the same linkage group with significant ( $P < 0.01$ ) and highly significant ( $p < 0.001$ ) linkage disequilibrium (LD) and  $R^2$  values at levels of chromosome and genome in durum wheat.**

Chromosome	No. of loci	Possible pairs	Observed pairs	Locus pairs with significant LD <sup>a</sup>				Mean $R^2$ (%)
				$P < 0.01$	$P < 0.001$	$R^2 > 0.1, P < 0.01$	$R^2 > 0.1, P < 0.001$	
1A	107	5671	2415	344 (6.07)	213 (3.76)	197 (3.47)	172 (3.03)	3.8
1B	97	4656	2775	514 (11.03)	341 (7.32)	292 (6.27)	271 (5.82)	4.1
2A	90	4005	2016	455 (11.36)	348 (8.69)	337 (8.41)	323 (8.07)	6.5
2B	82	3221	1711	348 (10.80)	244 (7.58)	212 (6.58)	198 (6.15)	5.6
3A	98	4753	2145	384 (8.08)	229 (4.82)	262 (5.51)	198 (4.17)	6.4
3B	66	2145	1176	225 (10.48)	164 (7.65)	142 (6.62)	136 (6.34)	4.9
4A	119	7021	3321	651 (9.27)	451 (6.42)	525 (7.48)	422 (6.01)	8.1
4B	72	2556	903	172 (6.73)	117 (4.58)	113 (4.42)	97 (3.80)	5.7
5A	83	3043	1653	271 (8.91)	206 (6.77)	195 (6.41)	183 (6.01)	5.3
5B	71	2485	946	208 (8.37)	158 (6.36)	146 (5.88)	134 (5.39)	5.9
6A	122	7381	2701	463 (6.27)	309 (4.19)	299 (4.05)	251 (3.40)	5.8
6B	103	5253	3240	673 (12.81)	421 (8.01)	392 (7.46)	346 (6.59)	6.4
7A	130	8385	3655	798 (9.52)	601 (7.17)	575 (6.86)	550 (6.56)	6.9
7B	98	4753	2211	401 (8.44)	264 (5.55)	269 (5.66)	232 (4.88)	6.4
<b>A genome</b>	749	40259	17906	3366 (8.36)	2357 (5.86)	2390 (5.94)	2099 (5.21)	6.2
<b>B genome</b>	589	25069	12962	2541 (10.13)	1709 (6.82)	1566 (6.25)	1414 (5.64)	5.6
<b>Whole</b>	1338	65328	30868	5907 (9.04)	4066 (6.22)	3956 (6.06)	3513 (5.38)	6.0

<sup>a</sup> Number of locus pairs and percentage of all possible locus pairs showing significant LD at  $P < 0.01$ ,  $P < 0.001$ ,  $R^2 > 0.1$  &  $P < 0.01$ , and  $R^2 > 0.1$  &  $P < 0.001$ , respectively.

**Table 2. SNP locus pairs in different linkage stage with significant ( $P < 0.01$ ) and highly significant ( $p < 0.001$ ) linkage disequilibrium (LD) and  $R^2$  values on genome level in durum wheat.**

Genome	No. of loci	Linkage stage	Possible pairs	Observed pairs	Locus pairs with significant LD				Mean $R^2$ (%)
					$P < 0.01$ (%) <sup>c</sup>	$P < 0.001$ (%) <sup>d</sup>	$R^2 > 0.1$ (%) <sup>e</sup>	$R^2 > 0.1$ (%) <sup>f</sup>	
A	749	Linked <sup>a</sup>	40259	17906	3366 (8.36)	2357 (5.86)	2390 (5.94)	2099 (5.21)	6.2
A	749	Unlinked <sup>b</sup>	239867	106844	18170 (7.58)	12036 (5.01)	11900 (4.96)	10370 (4.32)	5.3
A	749	Total	280126	124750	21536 (7.69)	14393 (5.14)	14290 (5.10)	12469 (4.45)	5.4
B	589	Linked <sup>a</sup>	25069	12962	2541 (10.12)	1709 (6.82)	1566 (6.25)	1414 (5.64)	5.6
B	589	Unlinked <sup>b</sup>	148097	74191	13233 (8.94)	8483 (5.73)	7862 (5.31)	7030 (4.75)	4.8
B	589	Total	173166	87153	15774 (9.11)	10192 (5.87)	9428 (5.44)	8444 (4.88)	4.9
Whole	1338	Linked <sup>a</sup>	65328	30868	5907 (9.04)	4066 (6.22)	3956 (6.06)	3513 (5.38)	6.0
Whole	1338	Unlinked <sup>b</sup>	829125	390035	67996 (8.20)	44488 (5.37)	42779 (5.16)	37669 (4.54)	5.0
Whole	1338	Total	894453	420903	73903 (8.26)	48554 (5.43)	46735 (5.22)	41182 (4.60)	5.1

<sup>a</sup> Locus on the same linkage group.

<sup>b</sup> Locus from different linkage groups.

<sup>c</sup> Locus pairs and percentage of all possible locus pairs showing significant LD at  $P < 0.01$ .

<sup>d</sup> Locus pairs and percentage of all possible locus pairs showing significant LD at  $P < 0.001$ .

<sup>e</sup> Locus pairs and percentage of all possible locus pairs showing significant LD at  $R^2 > 0.1$ ;  $P < 0.01$  adjusted for locus pairs description.

<sup>f</sup> Locus pairs and percentage of all possible locus pairs showing significant LD at  $R^2 > 0.1$ ;  $P < 0.001$  adjusted for locus pairs description.

doi:10.1371/journal.pone.0130854.t002

There were 894,453 possible pair-wise loci in the matrix of 150 genotypes and 1,338 SNP markers. Of these locus-pairs, 5.43% showed significant LD ( $p < 0.001$ ) (Table 2). There were 2,145 (3B) to 8,385 (7A) possible locus pairs in the 14 chromosomes. The percentage of locus pairs showing significant LD ( $p < 0.001$ ) ranged from 3.76% (1A) to 8.01% (6B), respectively. The average  $R^2$  values varied from 0.038 (1A) to 0.081 (4A) among the 14 chromosomes (Table 1). A small percentage of significant locus pairs had  $R^2$  value  $> 0.1$  ( $p < 0.001$ ). On the average, the highly significant pairs ( $R^2 > 0.1$ ;  $p < 0.001$ ) were 251 per chromosome, ranging from 97 (4B) to 550 (7A). The percentage of all possible locus pairs showing highly significant LD ( $R^2 > 0.1$ ;  $p < 0.001$ ) ranged from 3.03% (1A) to 6.59% (6B) (Table 1). The extent of LD was varying with chromosomes.

Table 2 showed LD value versus genetic distance in the locus pairs on genome level. There were 749 and 589 loci available for LD evaluations in the A and B genome, respectively. Across all 1,338 loci, 65,328 possible pairs of linked loci (in the same linkage groups) and 829,125 pairs of unlinked loci (from different linkage groups) were detected. The observed locus pairs of linked and unlinked loci were 30,868 and 390,935, respectively. Among the linked locus pairs, 2,357 (5.86%) possessed significant LD ( $P < 0.001$ ) in genome A, whereas, 1,709 (6.82%) had significant LD in genome B. As to the unlinked locus pairs, 1,236 (5.01%) had significant LD ( $p < 0.001$ ) in genome A, whereas 8,483 (5.73%) in the B genome.

The mean  $R^2$  values for all the linked pairs in genome A and B were 0.062 and 0.056, respectively. Therefore, the number of possible pairs, number of significant pairs, and mean  $R^2$  of the genome A were larger than the genome B except for the percentage of significant pairs (Table 2). The extent of LD was varying with chromosomes. The percentage of significant LD ( $R^2 > 0.1$ ;  $p < 0.001$ ) pairs in the A chromosomes generally was higher than the corresponding B



chromosomes except for 1A vs. 1B and 6A vs. 6B. The mean  $R^2$  value of the A chromosomes was higher than corresponding B chromosomes except for 1A vs. 1B; 5A vs. 5B and 6A vs. 6B (Tables 1 and 2). Thus the extent of LD of A genome was larger than the B genome on both the chromosome and genome levels in general.

### Variation of the key traits

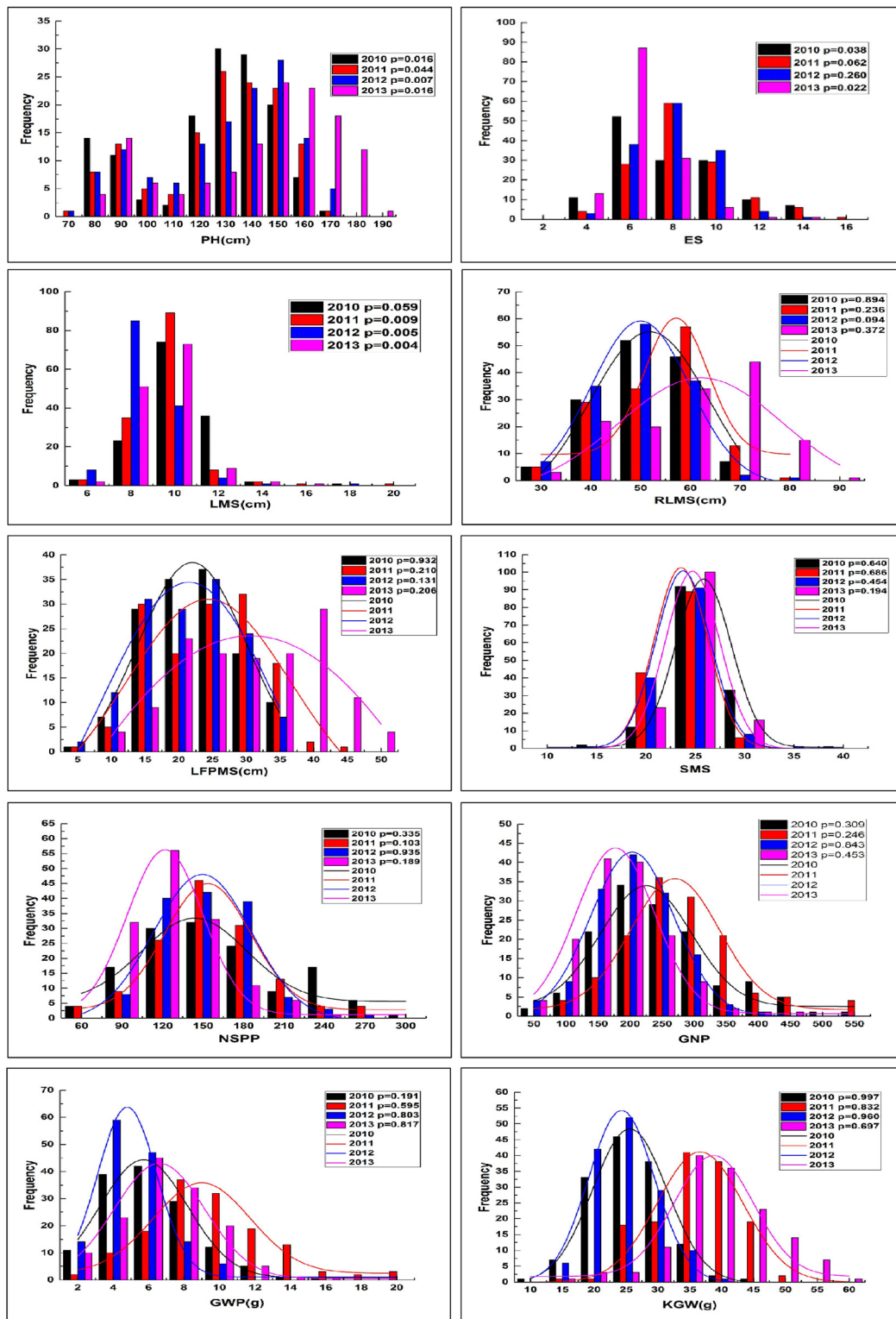
**Features of the examined traits.** All the durum accessions were observed for 10 agronomic and morphological traits in replicated field trials for four consecutive years (Table 3). Distribution histograms of the 10 traits were showed in Fig 1. In general, Kolmogorov-Smirnov test showed that most of the observed traits fitted the normal distribution except for PH, ES and LMS. PH significantly deviated from the normal distribution ( $P < 0.05$  in all the 4 years) and showed the feature of binomial distribution. ES significantly deviated from the normal distribution in 2010 and 2013 ( $P < 0.05$ ), and nearly significant in 2011 ( $P = 0.062$ ). LMS showed significant deviation ( $P < 0.05$ ) in 2011–2013 and nearly significant deviation in 2010 ( $P = 0.059$ ) (Fig 1). Therefore, most of the observed traits are quantitatively inherited. But PH seems controlled by a single gene together with polygene of minor effects in the population, and distribution of ES and LMS seems varying with the environment.

**Trait variation with year and genotype.** The trait distribution pattern was similar over the four years, and most of the traits generally showed normal distribution. The year effect was highly significant for most of the observed traits as revealed by the analysis of variance (ANOVA). The genotypic variation was highly significant for all the 10 traits. The genotype  $\times$  year ( $G \times E$ ) interaction effect was also highly significant for all the examined traits. Estimation of broad-sense heritability ( $H^2$ ) showed that most of the traits (6/10) have high heritability ( $H^2 > 65\%$ ) (Table 4). Therefore it is meaningful to conduct association analyses between the traits and SNP markers.

**Table 3. Mean values and variation of the 10 examined traits in four consecutive years.**

Year	Item	Trait <sup>a</sup>									
		PH	ES	LMS	RLMS	LFPMS	SMS	NSPP	GNP	GWP	KGW
2010	Average	121.68	6.88	9.25	46.36	19.46	22.97	144.99	223.21	5.28	23.30
	CV (%)	19.19	37.47	19.64	19.44	34.89	13.07	36.31	44.42	51.22	24.35
	Range	70–160.25	2.25–13.75	5.75–21.00	21–64.88	-3.50–33.50	14–36.00	55–273	21–562.75	0.2614.74	7.5–41.31
2011	Average	123.21	7.60	8.64	48.24	21.66	21.09	145.03	257.00	8.52	33.05
	CV (%)	19.56	33.29	17.24	21.10	35.68	11.89	33.46	37.45	40.88	19.82
	Range	63–160.83	2–22.67	4.88–18.15	24.60–71.00	3.10–41.33	14–28.50	42–386.00	56–650.33	0.83–19.48	13.75–47.85
2012	Average	125.00	7.08	7.77	44.26	18.93	21.17	136.97	179.39	4.18	22.40
	CV (%)	20.72	23.21	17.65	20.57	35.51	11.83	24.31	36.24	44.64	20.83
	Range	65.36–167.48	3.33–12.50	4.96–16.55	21.53–71.48	4.37–33.03	15.17–30.17	63.78–253.33	38.33–365.42	0.49–9.53	12.62–38.35
2013	Average	137.18	5.41	8.39	55.23	27.79	22.23	114.86	159.36	5.84	36.84
	CV (%)	21.90	26.55	21.05	23.34	35.97	11.43	27.52	41.54	43.68	21.19
	Range	61–183.58	3–12	5.65–22.13	25.25–81.1	5.75–49.4	16.5–29.33	61.5–228.5	13.75–404.75	0.44–14.67	10.19–55.42

<sup>a</sup> PH, plant height (cm); ES, number of effective spikes, LMS, length of main spike (cm); RLMS, rachis internode length of main spike (cm); LFPMS, panicle neck length of main spike (cm); SMS, spikelets on main spike; NSPP, number of spikelets per plant; GNP, Grain number per plant; GWP, grain weight per plant (g); KGW, 1000-grain weight (g).



**Fig 1. Frequency distribution of the 10 examined agronomic traits of durum wheat in four consecutive years.** P value of Kolmogorov-Smirnov test for each year was shown, the hypothesis of normal distribution could be accepted when  $P > 0.05$  (significant at  $P = 0.05$ ), and the trend lines of the accepted normal distribution were shown. PH, plant height (cm); ES, number of effective spikes; LMS, length of main spike (cm); RLMS, rachis internode length of main spike (cm); LFPMS, panicle neck length of main spike (cm); SMS, number of spikelets on main spike; NSPP, number of spikelets per plant; GNP, grain number per plant; GWP, grain weight per plant (g); KGW, 1000-grain weight (g).

doi:10.1371/journal.pone.0130854.g001



Table 4. Analysis of variance and heritability ( $H^2$ ) of the 10 examined traits.

Trait	Source of variation <sup>a</sup>	d.f.	MS	F	$H^2$ (%)
PH	Y	3	20205.29	411.19**	97.2
	G	141	7407.96	150.76**	
	Y * G	423	207.18	4.22**	
	E	1135	49.14		
	T	1703	16682.25		
ES	Y	3	358.6	83.87**	57.4
	G	141	23.98	5.61**	
	Y * G	423	10.21	2.39**	
	E	1135	4.28		
	T	1703	53.29		
LMS	Y	3	155.78	217.61**	91.7
	G	141	24.16	33.74**	
	Y * G	423	2	2.80**	
	E	1135	0.72		
	T	1703	75.36		
RLMS	Y	3	9334.72	419.17**	93.4
	G	141	1083.85	48.67**	
	Y * G	423	72.07	3.24**	
	E	1135	22.27		
	T	1703	2469.44		
LFPMS	Y	3	6863.44	396.37**	88.5
	G	141	568.42	32.83**	
	Y * G	423	65.22	3.77**	
	E	1135	17.32		
	T	1703	561.28		
SMS	Y	3	315.56	124.38**	87.9
	G	141	60.42	23.81**	
	Y * G	423	7.33	2.89**	
	E	1135	2.54		
	T	1703	485.34		
NSPP	Y	3	81443.16	45.01**	48.5
	G	141	8502.39	4.70**	
	Y * G	423	4377.13	2.42**	
	E	1135	1809.57		
	T	1703	21317.87		
GNP	Y	3	749025.54	94.43**	56.6
	G	141	36798.2	4.64**	
	Y * G	423	15976.96	2.01**	
	E	1135	7932.05		
	T	1703	55665.68		
GWP	Y	3	1423.18	138.19**	56
	G	141	38.15	3.70**	
	Y * G	423	16.78	1.63**	
	E	1135	10.3		
	T	1703	52.12		

(Continued)

Table 4. (Continued)

Trait	Source of variation <sup>a</sup>	d.f.	MS	F	H <sup>2</sup> (%)
KGW	Y	3	19924.11	664.88**	66.4
	G	141	235.22	7.85**	
	Y * G	423	79.04	2.64**	
	E	1135	29.97		
	T	1703	908.17		

\*\* significant at the probability level of 0.01.

<sup>a</sup> Y, years; G, genotype; E, error; T, total; d.f., degrees of freedom; MS, mean square; H<sup>2</sup>: broad-sense heritability.

doi:10.1371/journal.pone.0130854.t004

**Correlation among the observed traits.** Table 5 showed correlation coefficients among the 10 observed traits. Out of the 45 possible correlation pairs, more than 75% (34) were significant or highly significant. LMS, RLMS, LFPMS and SMS showed highly significant positive correlations with PH. NSPP, GNP and GWP showed highly significant positive correlations with ES, while SMS and KGW showed significant and highly significant negative correlations with ES. LMS showed significant positive correlations with RLMS and NSPP. The correlations between LFPMS and GNP, GWP were positive and highly significant. SMS was highly and positively correlated with NSPP, while negatively correlated with KGW. This indicated that the more SMS, the more NSPP correspondingly. In another word, the growth condition of main spike reflected the growth condition of the other spikes to some extent. And the more SMS and NSPP mean lighter and smaller grains. As a result, KGW was negatively correlated with SMS (Table 5).

### Association analysis

Association analyses between SNP markers and the 10 quantitative traits (PH, ES, LMS, SMS, RLMS, LFPMS, NSPP, GNP, GWP, and KGW) were conducted preliminarily under the GLM and MLM models by using the computer software TASSEL 3.0.124. Comparison between these two models showed that MLM decreased the total number of significant associations (p<0.01) (data not shown), and most of the significant associations were consistent between the two models. Yu and Buckler [34] suggested incorporating the pair-wise kinship (K matrix) as

Table 5. Correlation coefficients among the 10 observed agronomic traits.

Trait	PH	ES	LMS	RLMS	LFPMS	SMS	NSPP	GNP	GWP	KGW
PH	1									
ES	0.056	1								
LMS	0.242**	0.195**	1							
RLMS	0.848**	-0.015	0.133**	1						
LFLMS	0.674**	-0.012	-0.012	0.924**	1					
SMS	0.286**	-0.114*	0.326**	0.184**	0.044	1				
NSPP	0.169**	0.911**	0.290**	0.073	0.029	0.238**	1			
GNP	0.082	0.749**	0.219**	0.112*	0.125**	0.085	0.774**	1		
GWP	0.140**	0.534**	0.162**	0.272**	0.300**	-0.005	0.526**	0.832**	1	
KGW	0.156**	-0.145**	0.028	0.364**	0.407**	-0.095*	-0.167**	0.026	0.519**	1

\*, \*\* significant at the probability level of 0.05 and 0.01, respectively.

doi:10.1371/journal.pone.0130854.t005

random effects into a mixed model to correct relatedness and reduce the number of false positives in association analysis. In addition, association analyses in Yang et al [38] and Zhu and Yu [54] indicated that MLM (K+Q) model was better for correcting false positives associations than GLM. Therefore, the results under the MLM model that accounted for both Q and K matrixes were presented in this paper.

Some imperfect markers were excluded out of the 1,536 SNP markers. Thus 1,366 SNPs were used for association analysis in this study. Table 6 and S2 Table showed an overview and details of trait-marker associations under MLM model in four consecutive years, respectively. Fig 2 is the chromosome bin map showing candidate QTLs anchored by the associated SNP markers in durum wheat. In total, 201 significant associations were detected in the four years (60, 26, 45 and 70 for the year 2010, 2011, 2012 and 2013, respectively). The associations between SNP markers and traits were varying with the years.

In 2010, sixty markers were significantly associated with the ten observed traits. The distributions of the association pairs were uneven among the traits. Most of the associations were detected between markers and the yield traits. More than half of the markers were associated with GNP, and the number of associated markers for other traits range from 1 (ES, LFPMS and RLMS) to 10 (LMS). The percentage of the variation explained by marker ranged from 5.4% (CD454448\_6\_A\_84 associated with KGW) to 18.2% (BG605368\_2\_A\_Y\_310 associated with LMS).

In 2011, we detected 26 marker-trait association pairs. The number of the associated markers ranged from 1 (ES and NSPP) to 6 (LMS and GWP) (Table 6). The percentage of the variation explained by marker was in a range between 5.4% (BG274294\_1\_B\_382 associated with SMS) and 13.1% (BG605368\_2\_A\_Y\_310 associated with LMS).

In 2012, 45 marker-trait associations were detected. The number of the associated markers ranged from 2 (NSPP) to 10 (GNP). The percentage of the variation explained by marker varied from 5.2% (BG312827\_6\_A\_Y\_305 associated with PH) to 11.6% (BM134437\_3\_A\_Y\_233 associated with LMS).

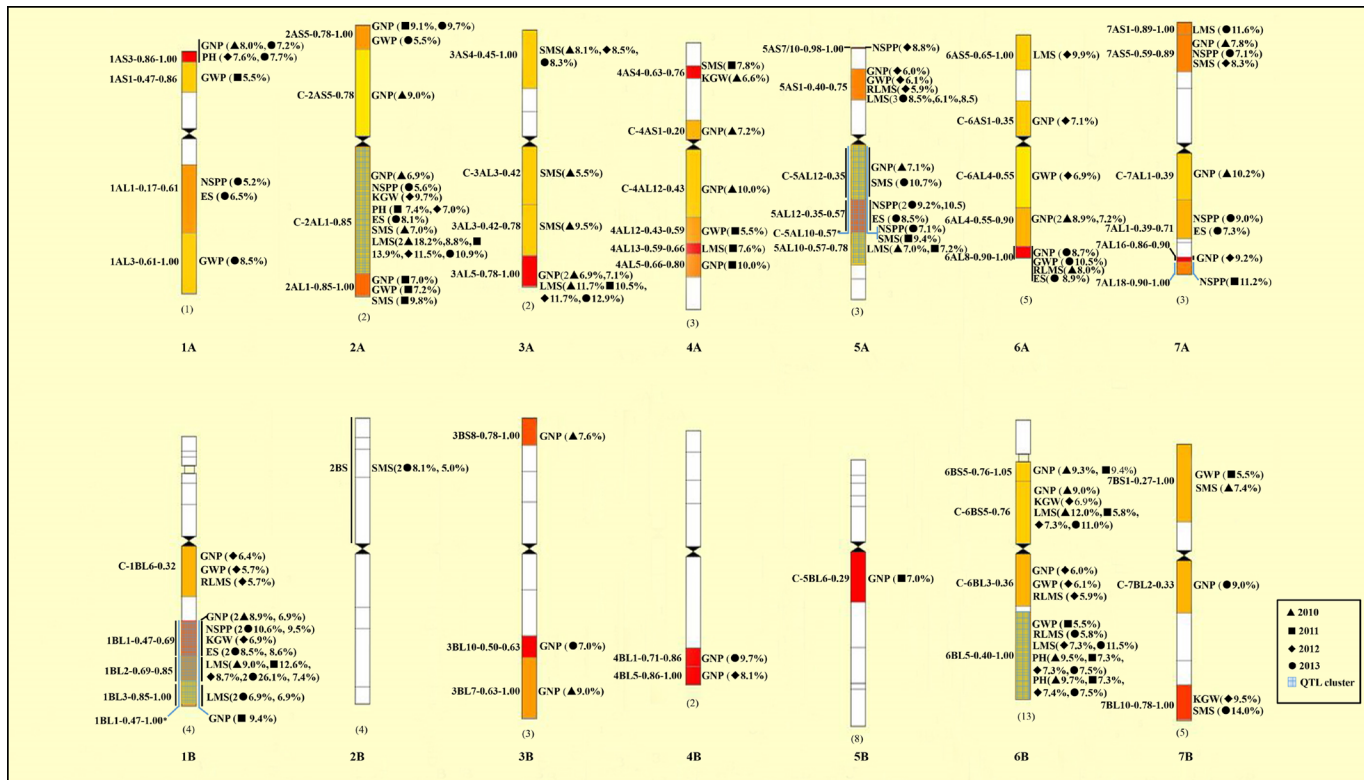
For the year 2013, 70 associations were detected. The percentage of the total variation explained by marker varied from 5.0% (BE444144\_2\_B\_N\_138 associated with SMS) to 26.1% (BF474284\_1\_B\_Y\_357 associated with LMS) (Table 6, S2 Table).

Moreover, taking consideration of all the four years, the number of markers associated with each trait ranged from 1 (LFPMS) to 54 (GNP), and the percentage of the total variation

**Table 6. Number of associated SNP markers in different years for the examined traits.**

Trait	Number of associated SNP markers				Total
	2010	2011	2012	2013	
PH	4	4	6	5	19
ES	0	1	0	12	13
LMS	10	6	9	15	40
RLMS	1	0	3	1	5
LFPMS	1	0	0	0	1
SMS	8	3	4	9	24
NSPP	0	1	2	13	16
GNP	33	5	10	10	58
GWP	0	6	7	5	18
KGW	3	0	4	0	7
<b>Total</b>	<b>60</b>	<b>26</b>	<b>45</b>	<b>70</b>	<b>201</b>

doi:10.1371/journal.pone.0130854.t006



**Fig 2. Chromosome bin map of plausible QTLs anchored by SNP markers in durum wheat.** The relative interval length is indicated on the left of each chromosome and QTLs represented by SNP-based associations and relative R value (%) are shown on the right. The number in front of the symbol means the repeats of the associations anchored in the interval in the corresponding years and without a number in front of the symbol means one repeat of the association anchored in the interval in one corresponding year. Details of the associations are presented in [S2 Table](#). The exact bins of some associated EST markers are unknown, and thus are shown below the chromosome.

doi:10.1371/journal.pone.0130854.g002

explained by marker ranged from 5.0% (*BE444144\_2\_B\_N\_138* associated with SMS) to 26.1% (*BF474284\_1\_B\_Y\_357* associated with LMS) ([S2 Table](#)). We found that one trait associated with many markers (e.g., GNP with 54 markers), and single markers were associated with multiple traits (*BE590553\_7\_A\_190* associated with GNP, NSPP and SMS, and *BE443538\_5\_A\_1436*, *BE590521\_6\_B\_N\_331* associated with GNP, GWP and RLMS, etc.). This may indicate that quantitative traits are always conferred by multiple loci, and QTLs conferring multiple agronomic traits may cluster around the single regions/markers due to pleiotropic effects of genes [55]. Seven associations (4 for LMS, 3 for PH) were detected in all the four years. Two associations (1 for PH, and 1 for SMS) were detected in three of the four years. Eleven associations were detected in two of the four years ([S2 Table](#)). These reproducible associations were significant and more reliable.

**Associations for morphological traits.** Plant height (PH): six significantly associated SNPs were detected in four years of 2010–2013 ([Table 6](#)). Three SNP markers, *BE405269\_4\_B\_84*, *BF475120\_6\_B\_67*, and *BF475120\_6\_B\_Y\_75* were detected to be significantly associated with PH in all the four years. Other three SNPs, *BG312827\_6\_A\_Y\_305*, *BE443948\_2\_A\_Y\_345* and *BE490041\_1\_A\_371* were significantly associated with PH in three or two of the four years ([S2 Table](#)). Furthermore, PH showed feature of the binomial distribution ([Fig 1](#)) and thus may be controlled by the polygene including a single major gene and some minor genes in the populations. These PH-associated SNP markers were mainly located in chromosome 1A,

2A, 4B, 6A and 6B. Several marker loci, significantly associated with PH, were previously detected on chromosomes 4B, 5A, 5B, 6B, 7A and 7B [56].

**RLMS and LFPMS:** A total of 5 and 1 SNP markers were detected in the four years for RLMS and LFPMS, respectively (Table 6, S2 Table). Markers significantly associated with the traits were present on chromosome 1B, 5A, 6A and 6B. *BE443538\_5\_A\_1436*, *BE590521\_6\_B\_N\_331* and *BG314205\_1\_B\_33* were associated with RLMS, GNP and GWP. Correlation analysis indicated significant positive correlations of RLMS with GNP and GWP (Table 5). Flag leaf and rachis internode were related to photosynthesis and photosynthetic product accumulation and transfer, and thus played important roles in grain filling process [57]. Therefore, it is understandable that SNP markers associated with RLMS and LFPMS are also related with GNP and GWP.

**LMS:** Six to fifteen associations were detected between LMS and SNP markers in the four years (Table 6, S2 Table). The SNP markers associated with LMS were located on chromosome 1B, 2A, 3A, 4A, 5A, 6A, 7A and 6B. Four SNP markers *BE445667\_6\_B\_Y\_285*, *BF474284\_1\_B\_Y\_357*, *BG605368\_2\_A\_Y\_310* and *BM134437\_3\_A\_Y\_233*, were significantly associated with LMS in all the four years. Five SNPs showed significant associations with LMS in two of the four years (S2 Table). The marker *BF484028\_5\_A\_Y\_97* corresponding to the *Vrn-A1* region in the interval of 5AL10-0.57–0.78 was significantly associated with LMS. Some associations were founded to be located in the same regions for LMS-related traits (GNP and GWP etc.) (Table 5, Fig 2).

**Associations for yield traits.** **ES and NSPP:** A total of 13 and 16 SNP markers were associated with ES and NSPP in the four years, respectively (Table 6, S2 Table). Some SNP markers were associated with both ES and NSPP. Highly significant positive correlation was detected between ES and NSPP (Table 5).

**SMS, GNP and GWP:** A total of 22, 54 and 18 significant associations with SNP markers were detected for SMS, GNP and GWP in the four years, respectively (Table 6, S2 Table). *BG314551\_3\_A\_Y\_162* was significantly associated with SMS in three of the four years. This SNP explained over 8.1% of the variation (Table 6, S2 Table). The EST represented by *BG314551\_3\_A\_Y\_162* was located in the same region as *Eps* gene (earliness per se). GWP showed positive correlation with GNP. Several SNP markers are thus associated with both GNP and GWP.

**KGW:** A total of 7 significant associations between KGW and SNP markers were detected in all the four years. These SNP markers associated with KGW were located in chromosomes 1B, 2A, 4A, 5B, 6A, 6B and 7B ( $R^2 = 4.9\text{--}9.7\%$ ), and mainly located in chromosomes 2A, 5B, 6A, 7A and 7B with  $R^2 > 9.2\%$  (S2 Table). Peng et al. [55] found eight QTLs for GWH (100-grain weight) on chromosomes 1B, 2A, 4A, 5A, 5B, 6B, 7A, and 7B, and major GWH QTLs were located on chromosomes 2A, 4A, and 5B. The marker *AY244508\_5\_B\_Y\_26*, significantly associated with KGW and GNP, was located in the same region as *API* and *Vrn-B1*.

## Discussion

### Linkage disequilibrium in durum wheat

The variation patterns of LD at both the chromosome and genome levels reflect the complicated evolutionary and breeding history in wheat [58]. In the present study, we demonstrated an extensive amount of LD in durum wheat using 1,338 SNP markers (Tables 1 and 2).

The extent of LD in A genome is higher than in B genome in general. The similar result was reported in previous study [59]. In their study based on SSR markers, the highest extent of significant LD was observed in D genome, followed by the A and B genomes of the bread wheat [59].



The genomic locations of genes controlling important adaptive traits were different. These can have a differential influence on LD in different genomes. *Vrn-A1* gene on chromosome 5A has higher number of widely distributed haplotypes than the *Vrn-B1* gene on chromosome 5B and thus more likely to have a stronger effect on LD [60]. In our study, chromosome 4B had the lowest percentage of significant LD pairs and mean  $R^2$  value, and thus possessed relatively low LD extent in chromosome 4B (Table 1). Akhunov et al. [61] also reported that chromosome 4B had the lowest number of haplotypes per locus and lowest haplotype diversity. This may indicate that the haplotype diversity and genes controlling important adaptive traits have a differential influence on LD in chromosome 4B. Therefore, the divergence in the extent of LD is probably related to breeding history and selection pressure applied to genes located in the different chromosomes and genomes during the process of cultivation [62].

The genetic diversity of genome A is lower than genome B [43, 55]. The extant LD in genome A is higher than in genome B, on the contrary. On chromosome level, some chromosomes have the similar extant LD (like 2A and 2B, 3A and 3B, 4A and 4B etc.) (Table 1). Chao et al [62] reported similar result. The extant LD was related to genetic diversity in the individual breeding program. The domestication history of genome A is longer than genome B in wheat [55, 63]. Genome A thus probably has more genes controlling important adaptive traits. Under the natural and artificial selections in the breeding programs, the genome A of cultivars captured comparable number of adaptive traits/genes, and widely distributed haplotypes resulting from the high extant LD [62, 63]. As mentioned above, breeding/domestication history and selection specific to each breeding program have influence on LD to some extent.

## Candidate QTLs revealed by association analysis

In the present study we performed association analysis using big number of SNP markers in durum wheat consisting of worldwide accessions. A total of 201 association pairs between SNP markers and 10 quantitative traits were detected in the four years (S2 Table). Fifty-two known regions were marked on the 14 chromosomes (Fig 2), which may represent the candidate QTLs.

Four credible SNP associations for PH were reproducible at least in three of the four consecutive years. These associations were located on 4B, 6A and 6B. Two markers (*BF475120\_6\_B\_67* and *BF475120\_6\_B\_Y\_75*) located on the same position in the region 6BL5-0.40–1.00 of the long arm of chromosome 6B, were associated with PH in all of the four years, and these two associations possibly represent a single credible QTL explaining over 7.2% of the variation in the four years (S2 Table). Several QTLs were reported in the similar region of 6BL by Börner et al. [64] and Cadalen et al. [56].

Four credible associations for LMS were reproducible in the four consecutive years. These associations were located on 1B, 2A, 3A and 6B, respectively, and thus might represent 4 QTLs. *BG605368\_2\_A\_Y\_310*, located on 2AL, was associated with LMS and explained 10.8% of the variation in the four years (S2 Table). Similar QTL for LMS was detected in the region of 2AL using SSR and EST-SSR markers in Yao et al [52], and Peng et al. [55] mapped over ten QTLs involving similar traits (PH, GNP, KGW and LMS) and defined two domestication factors in this chromosome arm. *BE445667\_6\_B\_Y\_285*, located on 6BL, was associated with LMS in the four years (S2 Table). QTLs involving similar traits (PH, GNP, KGW and LMS) were detected also in this region by Börner et al. [64].

The credible candidate QTLs may reside in a region containing several candidate genes conferring the examined traits. The candidate genes may have pleiotropic effects or several genes are clustered in the same region and acting on different traits [55]. Therefore, the candidate QTLs or the QTL-carried regions are potential reference regions for gene cluster. These QTLs

and the clustering regions are worthy of further precisely QTL locating and gene detecting and cloning.

## QTL clusters in the genome

As shown in Fig 2, most of the SNP associations were located on chromosomes 2A, 5A, 1B and 6B. The number of association effects in the A genome was larger than that in the B genome (Table 1, Fig 2). The genome A has longer domestication evolution history than the genome B in wheat, and thus probably has more genes controlling important adaptive traits [1, 55]. Chao et al. [62] demonstrated that the genome A of wheat cultivars captured comparable number of adaptive trait genes under the natural and artificial selection and in the breeding programs.

It is noteworthy that several associations co-locate in the same chromosome regions, even for the unrelated traits. There are several regions with association clusters especially on chromosomes 2A, 5A, 6A, 7A, 1B and 6B. For example, seven associations for PH, GNP, KGW and LMS are located on the proximal region C-2AL1-0.85 of chromosome 2 (S2 Table, Fig 2). Peng et al. [55] mapped over ten QTLs involving similar traits (PH, GNP, KGW and LMS) and defined two domestication factors in this chromosome arm. Yao et al. [52] detected similar QTLs for spike length, thousand kernel weight and spike number per plant in the same region. This region may be a convincing region for cluster of QTLs.

On the chromosome 5A, we detected association clusters for LMS, GNP, GWP and SMS mainly in the short arm (5AS1-0.40–0.75) and the long arm (5AL12-0.35–0.78) (Fig 2). Kato et al. [65] and Gadaleta et al. [66] reported QTL clusters for yield components (thousand kernel weight, grain yield per spike and kernel number per spike) in similar region 5AL15-0.67–0.78. Peng et al. [55] mapped 19 QTLs involving 11 traits including LMS, GNP, GWP and SMS and also defined two domestication factors in this chromosome 5AL arm. In Gadaleta et al. [66], many SNPs mapped in the bin 5AS1-0.40–0.75 on the short arm have duplicated loci in bin 5AL5-0.46–0.55 on the long arm. The bin on 5AS may have undergone a duplication followed by an insertion into the 5AL of the same chromosome 5A. This may explain the similar associations mapped in the regions of 5AS1-0.40–0.75 and 5AL12-0.35–0.78 (Fig 2).

Another significant cluster of associations for PH, GNP, KGW and LMS was detected on the long arm of chromosome 1B (1BL1-0.47–1.00) (Fig 2). Similarly, Börner et al. [64] detected QTLs for spike length and grain weight in this region. Similar result was reported by Cadalen et al. [56]. Peng et al. [55] mapped 8 QTLs involving 8 traits including LMS, GNP, GWP and SMS and defined one domestication factor in this 1BL chromosome arm.

Phenomenon of QTL clustering was formally reported by Peng et al. [55] for domestication-related traits in wild emmer wheat. They defined a cluster of QTLs co-located in the same chromosome region as domestication syndrome factor [55]. Actually this phenomenon of QTL clustering was repeatedly observed, although not verbally using the term of 'QTL cluster', in wheat [52, 56, 64–68]. In the present study, we demonstrated obvious QTL clusters represented by SNP-based associations in durum wheat (Fig 2). More and more studies tend to show that genes often reside in the genome in clusters. This seems especially true for resistance genes and QTLs for quantitatively inherited traits. The genetic mechanism for this universal phenomenon is the pleiotropic effect of genes [55]. Nevertheless, the genomic regions of QTL clusters need further validation by fine mapping and cloning of QTLs or genes.

## Genes for plant height

Plant height (PH) is the key agronomic trait in wheat. We found six marker-trait associations for PH located on chromosomes 1A, 2A, 4B, 6A and 6B in four years. Each of the two markers, *BF475120\_6\_B\_67* and *BF475120\_6\_B\_Y\_75*, associated with pH explained >7.0% of variation

in four years (S2 Table). In the chromosome region 6BL5-0.40–1.00 of *BF475120* (<http://wheat.pw.usda.gov/GG2/index.shtml>), the SSR marker *Xfbb250-6B* was founded to be significantly associated with PH [56]. As shown in NCBI database (<http://www.ncbi.nlm.nih.gov/>), *BF475120* is an EST sequence fragment derived from wheat salt-stressed crown cDNA library. The encoded protein of *BF475120* has very high homology ( $E = 1e^{-53}$ ) with the protein GDSL esterase/lipase from *Aegilops tauschii*. One member of rice GDSL esterase family might be involved in lipid yield [69]. Esterase/lipase is involved in the entire process of plant growth and development. Furthermore, Börner et al [64] detected two QTLs for PH on the similar region 6BL5-0.40–1.00 of 6BL. Thus it is reasonable that *BF475120* is associated with PH.

The SNP marker *BG312827\_6\_A\_Y\_305* associated with PH explained >5.2% of variation in the four consecutive years. The EST *BG312827* was derived from *T. monococcum* early reproductive apex cDNA library (<http://www.ncbi.nlm.nih.gov/>). The encoded protein has very high homology ( $E = 1e^{-63}$ ) with the DNA replication licensing factor, a mcm5-A-like enzyme from *Brachypodium distachyon* (<http://www.ncbi.nlm.nih.gov/>). DNA replication licensing factor expressed in shoot apex and flower buds is essential to undergo a single round of replication initiation and elongation per cell cycle [70]. *Arabidopsis* *MCM2* to *MCM5* and *MCM7* genes contain E2F consensus sites in their promoters. Their transcripts are elevated in plants expressing E2FA/DPA which not only regulates the mitotic cell cycle progression but also plays a role in the endocycle. It is a prerequisite for normal plant development [70–72]. Therefore *BG312827* closely relates with apex cell division and growth, and thus undoubtedly associate with PH.

Additionally, the marker *BE405269\_4\_B\_84* without exact site, located on chromosome 4B, was associated with PH in all the four years. This reproducible significant association is reliable. *Rht-B1*, located on chromosomes 4BS, is known to have major effect on PH [73]. The marker *BE405269\_4\_B\_84* was located in the same chromosome with *Rht-B1*, while the exacted region and relations need to be further explored.

## Genes for length of main spike

For length of main spike (LMS), we found a total of 23 SNP associations located on chromosomes 1B, 2A, 3A, 4A 5A, 6A, 7A and 6B in the four years. These reproducible associations are significant and reliable. *BF484028\_5\_A\_Y\_97* associated with LMS (S2 Table), and was mapped in the interval of 5AL10-0.57–0.78 (<http://wheat.pw.usda.gov/GG2/index.shtml>). Two genes *Vrn-A1* and *Fr1*, are located in the same chromosome interval as *BF484028\_5\_A\_Y\_97* [74]. *Vrn-A1*, a member of *Vrn-1* genes, regulates flowering-time, an important criterion for regional adaptation and yield in all the cereal crops [75]. *Vrn-1* gene is associated with heading date, spike length and grain yield. *Vrn-A1* had a greater effect on spike length [75–77]. Furthermore, *Vrn-1* completely links to MADS-box gene *API* [78] which defines the pattern of where floral organs arise, as well as determines development of the floral meristem [79, 80]. Therefore, the gene marked by *BF484028\_5\_A\_Y\_97* may affect LMS through *Vrn-A1* gene regulating vernalization.

The marker *BF474284\_1\_B\_Y\_357* associated with LMS explained >8.6% of the variation in the four consecutive years. *BF474284* is an EST derived from wheat vernalized crown cDNA library. It has complete homology ( $E = 0.0$ ) with *TAVDAC2* gene located on the long arm of chromosome 1B in wheat (<http://www.ncbi.nlm.nih.gov/>). The *Tavdac* cDNAs express in meristematic tissues (floral tissues and embryos), regulate the mitochondrial functions during the period of floral development to embryo formation [81]. Therefore, *Tavdac* is indirectly related to floral development and embryo formation in some ways, e.g., regulating the mitochondrial functions. This explained why *BF474284* was associated with LMS to some extent.

## Gene for number of spikelets on main spike

For number of spikelets on main spike (SMS), we found a total of 22 significant associations in the four years. One reliable SNP marker *BG314551\_3\_A\_Y\_162*, significantly associated with SMS in three years, explained over 8.1% of the variation (S2 Table). This SNP was located in the bin 3AS4-0.45–1.00 on chromosome arm 3AS in the same region as *Eps* gene (earliness per se). This gene is usually responsible for the fine-tuning of wheat flowering time. RFLP markers linked with *Eps* explained significant variation of plant height, thousand kernel weight, kernel number per spike, and grain yield [82, 83]. Thus *BG314551\_3\_A\_Y\_162* represent a significant factor from early reproductive apex greatly impacting SMS.

## Candidate gene for grain number per plant

Grain number per plant (GNP) is a key yield component factor in wheat. A total of 54 significant SNP associations were detected for GNP in the four years. Several reliable QTLs could be suggested for this trait (Table 6, S2 Table). *BF293541\_4\_A\_Y\_88* is located in the bin 4AL5-0.66–0.80 on chromosome arm 4AL (<http://wheat.pw.usda.gov/GG2/index.shtml>). This region was associated with spike length, spikelets density, grain number per spike [84].

The EST of *BF202706\_4\_A\_Y\_466* derived from wheat pre-anthesis spike cDNA library was mapped to wheat deletion bin 4AL12-0.43 (<http://wheat.pw.usda.gov/GG2/index.shtml>). This region harbors QTLs for grain yield, grain filling rate, spike length and grain number/m<sup>2</sup> [64, 85].

The EST of *BE498418\_7\_A\_148* was also derived from pre-anthesis spike cDNA library and mapped on 7AL (C-7AL1-0.39). This EST has very high homology ( $E = 1e^{-104}$ ) with UDP-D-xylose epimerase 3 coded by *UXE3* gene from *UXE* gene family in *Hordeum vulgare* (<http://www.ncbi.nlm.nih.gov/>). The abundant transcript of *HvUXE* was possibly correlated to arabinoxylan deposition in cell walls in the starchy endosperm during grain development. There was a substantial increase in *HvUXE1* and *HvUXE3* mRNA levels at the differentiation stage of endosperm development [86, 87]. The chromosome region of *BE498418* was also proved to carry the QTL for grain weight [64]. This further confirms the association of *BE498418\_7\_A\_148* with GNP.

The EST of *BG263521\_2\_A\_61* mapped in chromosome bin C-2AS5-0.78 (<http://wheat.pw.usda.gov/GG2/index.shtml>), was also derived from wheat pre-anthesis spike cDNA library, and has very high homology ( $E = 2e^{-126}$ ) with putative serine/threonine-protein kinase *WNK1* (<http://www.ncbi.nlm.nih.gov/>). *WNK1* gene is member of *WNK* gene family, which involved in the regulation of flowering time in *Arabidopsis* [88]. Several QTLs for grain yield and kernel number per spike were detected within this region [89]. Therefore, the associations between *BG263521\_2\_A\_61* and GNP may be true. Gene marked by SNP *BG263521\_2\_A\_61* affects GNP by regulating flowering time just as *WNK* does.

## Candidate gene for the 1000-grain weight

The 1000-grain weight (KGW) is another key yield component factor. A total of 7 significant associations between KGW and SNP markers mainly located in chromosomes 2A, 5B, 6A, 7A and 7B with  $R^2 > 5.4\%$ , were detected in all the four consecutive years (S2 Table). The SNP marker *AY244508\_5\_B\_Y\_26*, significantly associated with KGW and GNP and explained over 11% of variation, was located in the same region as *API* and *Vrn-B1*. *API* defines the genesis pattern of floral organs, as well as determines development of the floral meristem [79, 80]. *WAPI*, a wheat *APETALA1* homolog, plays a core role in the phase transition from vegetative to reproductive growth [90, 91]. Therefore, associations of *AY244508\_5\_B\_Y\_26* with KGW and GNP may be attributed to the role of *API* and *VRN1*.

Furthermore, in the composite map of wheat chromosome 5B (<http://wheat.pw.usda.gov/GG2/index.shtml>), three QTLs (*QGpc.ndsu-5B.1*, *QYld.ndsu-5B* and *QGw1.inra-5B*) lie in the interval *Xmwwg922-Xcdo1326.1* affect KGW and grain yield around the *Vrn-B1* locus [92, 93]. Thus there may be many loci on chromosome 5B controlling grain weight.

*BG605368\_2\_A\_Y\_310* was associated with KGW, and explained 9.71% of variation (S2 Table). As discussed above, *BG605368\_2\_A\_Y\_310* was also associated with LMS in all the four years. The EST *BG605368* was derived from wheat pre-anthesis spike cDNA library. It is highly homologous ( $E = 1e^{-127}$ ) with Exopolygalacturonase from *T. urartu*. Exopolygalacturonase expressed in pollen and young developing tissues, suggesting that they could be implicated in the cell wall modifications and related to cell elongation and/or expansion in these tissues [94]. *BG605368* may be related to flower development. Several QTLs for grain weight and yield in the region (C-2AL1-0.85) of the EST were detected in previous study [64, 95, 96]. Therefore, the association between *BG605368\_2\_A\_Y\_310* and KGW and LMS should be credible.

## Conclusions

The previous studies indicated that both QTL analysis and association mapping are suitable and effective tools for mapping quantitative loci in wheat and barley [7, 9, 55, 97–99]. We detected 201 significant associations in total between SNP markers and 10 quantitative traits in durum wheat in four years. Some of the associations are corroborated by the previous QTL analyses, and further supported by the functions of the deriving ESTs and the homologous genes. The plausible QTLs represented by the associated SNP markers are generally clustered in specific chromosome regions of the wheat genome, especially 2A, 5A, 6A, 7A, 1B, and 6B chromosomes. Nevertheless, the associated SNP markers need to be further confirmed before they can be utilized in marker-assisted selection breeding programs [7, 9, 100].

## Supporting Information

**S1 Table. Durum wheat accessions used in the study.** Accession identifier, accession name, place of origin and year of collection are listed for each of the 150 entries.  
(DOCX)

**S2 Table. Significant trait-SNP marker pairs in four consecutive years.** <sup>a</sup> PH, plant height; ES, number of effective spikes; LMS, length of main spike; RLMS, rachis internode length of main spike; LFPMS, pillow neck length of main spike; SMS, spikelets on main spike; NSPP, number of spikelets per plant; GNP, Grain number per plant; GWP, grain weight per plant; KGW, 1000-grain weight; <sup>b</sup> P: the permutation based test for marker significance of individual markers; <sup>c</sup> R<sup>2</sup>: the fraction of the total variation explained by the marker after fitting the other model effects.  
(DOCX)

## Acknowledgments

We sincerely thank Ms. Robin Permut, the English editor at the Institute of Evolution, University of Haifa, for her professionally editing the English of this paper. We are also greatly indebted to the two anonymous reviewers for their critical, helpful and constructive comments on this manuscript.



## Author Contributions

Conceived and designed the experiments: DS JP. Performed the experiments: XH JR SH SS ML. Analyzed the data: XH JR XR ML. Contributed reagents/materials/analysis tools: ML EN JP DS. Wrote the paper: XH JP DS. Drafting the article or revising it critically for important intellectual content: XH JP DS XR CF.

## References

1. Peng J, Sun D, Nevo E. Wild emmer wheat, *Triticum dicoccoides*, occupies a pivotal position in wheat domestication process. *Aust J Crop Sci.* 2011; 5: 1127–1143.
2. Nachit M, Nachit G, Ketata H, Gauch H Jr, Zobel R. Use of AMMI and linear regression models to analyze genotype-environment interaction in durum wheat. *Theor Appl Genet.* 1992; 83: 597–601. doi: [10.1007/BF00226903](https://doi.org/10.1007/BF00226903) PMID: [24202676](https://pubmed.ncbi.nlm.nih.gov/24202676/)
3. De Vita P, Nicosia OLD, Nigro F, Platani C, Riefolo C, Di Fonzo N, et al. Breeding progress in morpho-physiological, agronomical and qualitative traits of durum wheat cultivars released in Italy during the 20th century. *Eur J of Agron.* 2007; 26: 39–53.
4. Kresovich S, Szewc-McFadden A, Bliet S, McFerson J. Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rape-seed). *Theor Appl Genet.* 1995; 91: 206–211. doi: [10.1007/BF00220879](https://doi.org/10.1007/BF00220879) PMID: [24169765](https://pubmed.ncbi.nlm.nih.gov/24169765/)
5. Nevo E, Beiles A. Genetic diversity of wild emmer wheat in Israel and Turkey. *Theor Appl Genet.* 1989; 77: 421–455. doi: [10.1007/BF00305839](https://doi.org/10.1007/BF00305839) PMID: [24232622](https://pubmed.ncbi.nlm.nih.gov/24232622/)
6. Jana S. Some recent issues on the conservation of crop genetic resources in developing countries. *Genome.* 1999; 42: 562–569.
7. Sun D, Ren W, Sun G, Peng J. Molecular diversity and association mapping of quantitative traits in Tibetan wild and worldwide originated barley (*Hordeum vulgare* L.) germplasm. *Euphytica.* 2011; 178: 31–43.
8. Matus I, Hayes P. Genetic diversity in three groups of barley germplasm assessed by simple sequence repeats. *Genome.* 2002; 45: 1095–1106. PMID: [12502254](https://pubmed.ncbi.nlm.nih.gov/12502254/)
9. Peng J, Bai Y, Haley S, Lapitan N. Microsatellite-based molecular diversity of bread wheat germplasm and association mapping of wheat resistance to the Russian wheat aphid. *Genetica.* 2009; 135: 95–122. doi: [10.1007/s10709-008-9262-x](https://doi.org/10.1007/s10709-008-9262-x) PMID: [18392559](https://pubmed.ncbi.nlm.nih.gov/18392559/)
10. Dograr N, Akin-Yalin S, Akkaya M. Discriminating durum wheat cultivars using highly polymorphic simple sequence repeat DNA markers. *Plant Breeding.* 2000; 119: 360–362.
11. Eujayl I, Sorrells M, Baum M, Wolters P, Powell W. Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theor Appl Genet.* 2002; 104: 399–407. PMID: [12582712](https://pubmed.ncbi.nlm.nih.gov/12582712/)
12. Incirli A, Akkaya MS. Assessment of genetic relationships in durum wheat cultivars using AFLP markers. *Genet Resour Crop Ev.* 2001; 48: 233–238.
13. Pujar S, Tamhankar S, Rao V, Gupta V, Naik S, Ranjekar P. Arbitrarily primed-PCR based diversity assessment reflects hierarchical groupings of Indian tetraploid wheat genotypes. *Theor Appl Genet.* 1999; 99: 868–876.
14. Soleimani V, Baum B, Johnson D. AFLP and pedigree-based genetic diversity estimates in modern cultivars of durum wheat [*Triticum turgidum* L. subsp. *durum* (Desf.) Husn.]. *Theor Appl Genet.* 2002; 104: 350–357. PMID: [12582707](https://pubmed.ncbi.nlm.nih.gov/12582707/)
15. Akhunov E, Nicolet C, Dvorak J. Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor Appl Genet.* 2009; 119: 507–517. doi: [10.1007/s00122-009-1059-5](https://doi.org/10.1007/s00122-009-1059-5) PMID: [19449174](https://pubmed.ncbi.nlm.nih.gov/19449174/)
16. Deschamps S, Campbell MA. Utilization of next-generation sequencing platforms in plant genomics and genetic variant discovery. *Mol Breeding.* 2010; 25: 553–570.
17. Trebbi D, Maccaferri M, de Heer P, Sørensen A, Giuliani S, Salvi S, et al. High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). *Theor Appl Genet.* 2011; 123: 555–569. doi: [10.1007/s00122-011-1607-7](https://doi.org/10.1007/s00122-011-1607-7) PMID: [21611761](https://pubmed.ncbi.nlm.nih.gov/21611761/)
18. Edwards KJ, Reid AL, Coghill JA, Berry ST, Barker GL. Multiplex single nucleotide polymorphism (SNP)-based genotyping in allohexaploid wheat using padlock probes. *Plant Biotechnol J.* 2009; 7: 375–390. doi: [10.1111/j.1467-7652.2009.00413.x](https://doi.org/10.1111/j.1467-7652.2009.00413.x) PMID: [19379286](https://pubmed.ncbi.nlm.nih.gov/19379286/)
19. Ganai MW, Altmann T, Röder MS. SNP identification in crop plants. *Curr Opin Plant Biol.* 2009; 12: 211–217. doi: [10.1016/j.pbi.2008.12.009](https://doi.org/10.1016/j.pbi.2008.12.009) PMID: [19186095](https://pubmed.ncbi.nlm.nih.gov/19186095/)

20. Varshney RK, Nayak SN, May GD, Jackson SA. Next-generation sequencing technologies and their implications for crop genetics and breeding. *Trends Biotechnol.* 2009; 27: 522–530. doi: [10.1016/j.tibtech.2009.05.006](https://doi.org/10.1016/j.tibtech.2009.05.006) PMID: [19679362](https://pubmed.ncbi.nlm.nih.gov/19679362/)
21. Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, et al. Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array. *Plant Biotechnol J.* 2014; 12: 787–796. doi: [10.1111/pbi.12183](https://doi.org/10.1111/pbi.12183) PMID: [24646323](https://pubmed.ncbi.nlm.nih.gov/24646323/)
22. Cho RJ, Mindrinos M, Richards DR, Sapolsky RJ, Anderson M, Drenkard E, et al. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nat Genet.* 1999; 23: 203–207. PMID: [10508518](https://pubmed.ncbi.nlm.nih.gov/10508518/)
23. Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, et al. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Res.* 2002; 9: 163–171. PMID: [12465716](https://pubmed.ncbi.nlm.nih.gov/12465716/)
24. Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, et al. A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. *Genetics.* 2007; 176: 685–696. PMID: [17339218](https://pubmed.ncbi.nlm.nih.gov/17339218/)
25. Kota R, Varshney R, Prasad M, Zhang H, Stein N, Graner A. EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. *Funct Integr Genomic.* 2008; 8: 223–233. PMID: [17968603](https://pubmed.ncbi.nlm.nih.gov/17968603/)
26. Chao S, Zhang W, Akhunov E, Sherman J, Ma Y, Luo MC, et al. Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. *Mol Breeding.* 2009; 23: 23–33.
27. Somers DJ, Kirkpatrick R, Moniwa M, Walsh A. Mining single-nucleotide polymorphisms from hexaploid wheat ESTs. *Genome.* 2003; 46: 431–437. PMID: [12834059](https://pubmed.ncbi.nlm.nih.gov/12834059/)
28. Kozlova S, Khlestkina E, Salina E. Specific features in using SNP markers developed for allopolyploid wheat. *Russ J Genet.* 2009; 45: 81–84.
29. Flint-Garcia SA, Thornsberry JM, Buckler ES. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol.* 2003; 54: 357–374. PMID: [14502995](https://pubmed.ncbi.nlm.nih.gov/14502995/)
30. Zondervan KT, Cardon LR. The complex interplay among factors that influence allelic association. *Nat Rev Genet.* 2004; 5: 89–100. PMID: [14735120](https://pubmed.ncbi.nlm.nih.gov/14735120/)
31. Bodmer WF. Human genetics: the molecular challenge. *BioEssays.* 1987; 7: 41–45. PMID: [3632655](https://pubmed.ncbi.nlm.nih.gov/3632655/)
32. Thomas DC, Haile RW, Duggan D. Recent developments in genomewide association scans: a workshop summary and review. *Ame J Hum Genet.* 2005; 77: 337–345. PMID: [16080110](https://pubmed.ncbi.nlm.nih.gov/16080110/)
33. Mackay I, Powell W. Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci.* 2007; 12: 57–63. PMID: [17224302](https://pubmed.ncbi.nlm.nih.gov/17224302/)
34. Yu J, Buckler ES. Genetic association mapping and genome organization of maize. *Curr Opin in Biotech.* 2006; 17: 155–160. PMID: [16504497](https://pubmed.ncbi.nlm.nih.gov/16504497/)
35. Risch NJ. Searching for genetic determinants in the new millennium. *Nature.* 2000; 405: 847–856. PMID: [10866211](https://pubmed.ncbi.nlm.nih.gov/10866211/)
36. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet.* 2010; 42: 961–967. doi: [10.1038/ng.695](https://doi.org/10.1038/ng.695) PMID: [20972439](https://pubmed.ncbi.nlm.nih.gov/20972439/)
37. Pasam RK, Sharma R, Malosetti M, van Eeuwijk FA, Haseneyer G, Kilian B, et al. Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biol.* 2012; 12: 16. doi: [10.1186/1471-2229-12-16](https://doi.org/10.1186/1471-2229-12-16) PMID: [22284310](https://pubmed.ncbi.nlm.nih.gov/22284310/)
38. Yang X, Yan J, Shah T, Warburton ML, Li Q, Li L, et al. Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. *Theor Appl Genet.* 2010; 121: 417–431. doi: [10.1007/s00122-010-1320-y](https://doi.org/10.1007/s00122-010-1320-y) PMID: [20349034](https://pubmed.ncbi.nlm.nih.gov/20349034/)
39. Buckler E, Gore M. An *Arabidopsis* haplotype map takes root. *Nat Genet.* 2007; 39: 1056–1057. PMID: [17728772](https://pubmed.ncbi.nlm.nih.gov/17728772/)
40. Zhu C, Gore M, Buckler ES, Yu J. Status and prospects of association mapping in plants. *Plant Genome.* 2008; 1: 5–20.
41. Carvalho A, Guedes-Pinto H, Lima-Brito JE. Genetic diversity in old Portuguese durum wheat cultivars assessed by retrotransposon-based markers. *Plant Mol Biol Rep.* 2012; 30: 578–589.
42. Laido G, Marone D, Russo MA, Colecchia SA, Mastrangelo AM, De Vita P, et al. Linkage disequilibrium and genome-wide association mapping in tetraploid wheat (*Triticum turgidum* L.). *PloS One.* 2014; 9: e95211. doi: [10.1371/journal.pone.0095211](https://doi.org/10.1371/journal.pone.0095211) PMID: [24759998](https://pubmed.ncbi.nlm.nih.gov/24759998/)
43. Ren J, Sun D, Chen L, You FM, Wang J, Peng Y, et al. Genetic diversity revealed by single nucleotide polymorphism markers in a worldwide germplasm collection of durum wheat. *Int J Mol Sci.* 2013; 14: 7061–7088. doi: [10.3390/ijms14047061](https://doi.org/10.3390/ijms14047061) PMID: [23538839](https://pubmed.ncbi.nlm.nih.gov/23538839/)
44. Stein N, Herren G, Keller B. A new DNA extraction method for high-throughput marker analysis in a large-genome species such as *Triticum aestivum*. *Plant Breeding.* 2001; 120: 354–356.

45. Belzile F, Somers DJ, Banks T, DePauw R, Fox S, Clarke J, et al. Genome-wide linkage disequilibrium analysis in bread wheat and durum wheat. *Genome*. 2007; 50:557–67. PMID: [17632577](#)
46. Rafalski A, Morgante M. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet*. 2004; 20: 103–111. PMID: [14746992](#)
47. Stich B, Melchinger AE, Frisch M, Maurer HP, Heckenberger M, Reif JC. Linkage disequilibrium in European elite maize germplasm investigated with SSRs. *Theor Appl Genet*. 2005; 111: 723–730. PMID: [15997389](#)
48. Hill W, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet*. 1968; 38: 226–231. doi: [10.1007/BF01245622](#) PMID: [24442307](#)
49. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155: 945–959. PMID: [10835412](#)
50. Lynch M, Ritland K. Estimation of pairwise relatedness with molecular markers. *Genetics*. 1999; 152: 1753–1766. PMID: [10430599](#)
51. Hardy OJ, Vekemans X. spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol Ecol Notes*. 2002; 2: 618–620.
52. Yao J, Wang L, Liu L, Zhao C, Zheng Y. Association mapping of agronomic traits on chromosome 2A of wheat. *Genetica*. 2009; 137: 67–75. doi: [10.1007/s10709-009-9351-5](#) PMID: [19160058](#)
53. Weir BS, Cockerham C. Genetic data analysis II: Methods for discrete population genetic data. Sinauer Assoc. Inc, Sunderland, MA, USA. 1996.
54. Zhu C, Yu J. Nonmetric multidimensional scaling corrects for population structure in association mapping with different sample types. *Genetics*. 2009; 182: 875–888. doi: [10.1534/genetics.108.098863](#) PMID: [19414565](#)
55. Peng J, Ronin Y, Fahima T, Röder MS, Li Y, Nevo E, et al. Domestication quantitative trait loci in *Triticum dicoccoides*, the progenitor of wheat. *P Natl Acad Sci USA*. 2003; 100: 2489–2494. PMID: [12604784](#)
56. Cadalen T, Sourdille P, Charmet G, Tixier M, Gay G, Boeuf C, et al. Molecular markers linked to genes affecting plant height in wheat using a doubled-haploid population. *Theor Appl Genet*. 1998; 96: 933–940.
57. Blum A. Photosynthesis and transpiration in leaves and ears of wheat and barley varieties. *J Exp Bot*. 1985; 36: 432–440.
58. Dubcovsky J, Dvorak J. Genome plasticity a key factor in the success of polyploid wheat under domestication. *Science*. 2007; 316: 1862–1866. PMID: [17600208](#)
59. Chen X, Min D, Yasir TA, Hu YG. Genetic Diversity, Population structure and linkage disequilibrium in elite Chinese winter wheat investigated with SSR markers. *PLoS One*. 2012; 7: e44510. doi: [10.1371/journal.pone.0044510](#) PMID: [22957076](#)
60. Zhang X, Xiao Y, Zhang Y, Xia X, Dubcovsky J, He Z. Allelic variation at the vernalization genes, and in Chinese wheat cultivars and their association with growth habit. *Crop Sci*. 2008; 48: 458–570.
61. Akhunov E, Akhunova A, Anderson O, Anderson J, Blake N, Clegg M, et al. Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC Genomics*. 2010; 11: 702. doi: [10.1186/1471-2164-11-702](#) PMID: [21156062](#)
62. Chao S, Dubcovsky J, Dvorak J, Luo MC, Baenziger SP, Matnyazov R, et al. Population- and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.). *BMC Genomics*. 2010; 11: 727. doi: [10.1186/1471-2164-11-727](#) PMID: [21190581](#)
63. Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, et al. Molecular basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). *Plant Cell*. 2005; 17: 1033–1045. PMID: [15749759](#)
64. Börner A, Schumann E, Fürste A, Cöster H, Leithold B, Röder M, et al. Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). *Theor Appl Genet*. 2002; 105: 921–936. PMID: [12582918](#)
65. Kato K, Miura H, Sawada S. Mapping QTLs controlling grain yield and its components on chromosome 5A of wheat. *Theor Appl Genet*. 2000; 101: 1114–1121.
66. Gadaleta A, Giancaspro A, Nigro D, Giove S, Incerti O, Simeone R, et al. A new genetic and deletion map of wheat chromosome 5A to detect candidate genes for quantitative traits. *Mol Breeding*. 2014; 34: 1599–1611.
67. Chantret N, Sourdille P, Röder M, Tavaud M, Bernard M, Doussinault G. Location and mapping of the powdery mildew resistance gene MIRE and detection of a resistance QTL by bulked segregant analysis (BSA) with microsatellites in wheat. *Theor Appl Genet*. 2000; 100: 1217–1224.

68. Sourdille P, Cadalen T, Guyomarch H, Snape J, Perretant M, Charmet G, et al. An update of the Courtot × Chinese Spring intervarietal molecular marker linkage map for the QTL detection of agronomic traits in wheat. *Theor Appl Genet*. 2003; 106: 530–538. PMID: [12589554](#)
69. Akoh CC, Lee GC, Liaw YC, Huang TH, Shaw JF. GDSL family of serine esterases/lipases. *Prog Lipid Res*. 2004; 43: 534–552. PMID: [15522763](#)
70. Shultz RW, Lee TJ, Allen GC, Thompson WF, Hanley-Bowdoin L. Dynamic localization of the DNA replication proteins *MCM5* and *MCM7* in plants. *Plant Physiol*. 2009; 150: 658–669. doi: [10.1104/pp.109.136614](#) PMID: [19357199](#)
71. Stevens R, Mariconti L, Rossignol P, Perennes C, Cella R, Bergounioux C. Two *E2F* sites in the *Arabidopsis MCM3* promoter have different roles in cell cycle activation and meristematic expression. *J Biol Chem*. 2002; 277: 32978–32984. PMID: [12089153](#)
72. Vandepoele K, Vlieghe K, Florquin K, Hennig L, Beemster GT, Gruissem W, et al. Genome-wide identification of potential plant *E2F* target genes. *Plant Physiol*. 2005; 139: 316–328. PMID: [16126853](#)
73. Börner A, Plaschke J, Korzun V, Worland A. The relationships between the dwarfing genes of wheat and rye. *Euphytica*. 1996; 89: 69–75.
74. Sutka J, Galiba G, Vagujfalvi A, Gill B, Snape J. Physical mapping of the *Vrn-A1* and *Fr1* genes on chromosome 5A of wheat using deletion lines. *Theor Appl Genet*. 1999; 99: 199–202.
75. Li W, Nelson J, Chu C, Shi L, Huang S, Liu D. Chromosomal locations and genetic relationships of tiller and spike characters in wheat. *Euphytica*. 2002; 125: 357–366.
76. Kato K, Miura H, Sawada S. QTL mapping of genes controlling ear emergence time and plant height on chromosome 5A of wheat. *Theor Appl Genet*. 1999; 98: 472–477.
77. Sun QM, Zhou RH, Gao LF, Zhao GY, Jia JZ. The characterization and geographical distribution of the genes responsible for vernalization requirement in Chinese bread wheat. *J Integr Plant Biol*. 2009; 51: 423–432. doi: [10.1111/j.1744-7909.2009.00812.x](#) PMID: [19341410](#)
78. Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J. Positional cloning of the wheat vernalization gene *VRN1*. *P Natl Acad Sci USA*. 2003; 100: 6263–6268. PMID: [12730378](#)
79. Irish VF, Sussex IM. Function of the *apetala-1* gene during *Arabidopsis* floral development. *Plant Cell*. 1990; 2: 741–753. PMID: [1983792](#)
80. Kaufmann K, Wellmer F, Muiño JM, Ferrer T, Wuest SE, Kumar V, et al. Orchestration of floral initiation by *APETALA1*. *Science*. 2010; 328: 85–89. doi: [10.1126/science.1185244](#) PMID: [20360106](#)
81. Elkeles A, Devos KM, Graur D, Zizi M, Breiman A. Multiple cDNAs of wheat voltage-dependent anion channels (VDAC): isolation, differential expression, mapping and evolution. *Plant Mol Biol*. 1995; 29: 109–124. PMID: [7579156](#)
82. Campbell B, Baenziger PS, Gill K, Eskridge KM, Budak H, Erayman M, et al. Identification of QTLs and environmental interactions associated with agronomic traits on chromosome 3A of wheat. *Crop Sci*. 2003; 43: 1493–1505.
83. Shah M, Gill K, Baenziger P, Yen Y, Kaeppler S, Ariyaratne H. Molecular mapping of loci for agronomic traits on chromosome 3A of bread wheat. *Crop Sci*. 1999; 39: 1728–1732.
84. Liu L, Wang L, Yao J, Zheng Y, Zhao C. Association mapping of six agronomic traits on chromosome 4A of wheat (*Triticum aestivum* L.). *Mol Plant Breeding*. 2010; 1: 1–10.
85. Kirigwi F, Van Ginkel M, Brown-Guedira G, Gill B, Paulsen G, Fritz A. Markers associated with a QTL for grain yield in wheat under drought. *Mol Breeding*. 2007; 20: 401–413.
86. Zhang Q, Shirley N, Lahnstein J, Fincher GB. Characterization and expression patterns of UDP-D-glucuronate decarboxylase genes in barley. *Plant Physiol*. 2005; 138: 131–141. PMID: [15849307](#)
87. Zhang Q, Shirley NJ, Burton RA, Lahnstein J, Hrmova M, Fincher GB. The genetics, transcriptional profiles, and catalytic properties of UDP- $\alpha$ -D-xylose 4-epimerases from barley. *Plant Physiol*. 2010; 153: 555–568. doi: [10.1104/pp.110.157644](#) PMID: [20435741](#)
88. Wang Y, Liu K, Liao H, Zhuang C, Ma H, Yan X. The plant *WNK* gene family and regulation of flowering time in *Arabidopsis*. *Plant Biology*. 2008; 10: 548–562. doi: [10.1111/j.1438-8677.2008.00072.x](#) PMID: [18761494](#)
89. Li S, Jia J, Wei X, Zhang X, Li L, Chen H, et al. A intervarietal genetic map and QTL analysis for yield traits in wheat. *Mol Breeding*. 2007; 20: 167–178.
90. Murai K, Miyamae M, Kato H, Takumi S, Ogihara Y. *WAP1*, a wheat *APETALA1* homolog, plays a central role in the phase transition from vegetative to reproductive growth. *Plant Cell Physiol*. 2003; 44: 1255–1265. PMID: [14701921](#)
91. Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES. MADS box genes control vernalization-induced flowering in cereals. *P Natl Acad Sci USA*. 2003; 100: 13099–13104. PMID: [14557548](#)

92. Gonzalez-Hernandez J, Elias E, Kianian S. Mapping genes for grain protein concentration and grain yield on chromosome 5B of *Triticum turgidum* (L.) var. *dicoccoides*. *Euphytica*. 2004; 139: 217–225.
93. Groos C, Robert N, Bervas E, Charmet G. Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. *Theor Appl Genet*. 2003; 106: 1032–1040. PMID: [12671751](#)
94. Torki M, Mandaron P, Thomas F, Quigley F, Mache R, Falconet D. Differential expression of a polygalacturonase gene family in *Arabidopsis thaliana*. *Mol Gen Genet*. 1999; 261: 948–952. PMID: [10485285](#)
95. Huang X, Cöster H, Ganai M, Röder M. Advanced backcross QTL analysis for the identification of quantitative trait loci alleles from wild relatives of wheat (*Triticum aestivum* L.). *Theor Appl Genet*. 2003; 106: 1379–1389. PMID: [12750781](#)
96. McCartney C, Somers D, Humphreys D, Lukow O, Ames N, Noll J, et al. Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452×'AC Domain'. *Genome*. 2005; 48: 870–883. PMID: [16391693](#)
97. Cockram J, White J, Leigh FJ, Lea VJ, Chiapparino E, Laurie DA, et al. Association mapping of partitioning loci in barley. *BMC Genet*. 2008; 9: 16. doi: [10.1186/1471-2156-9-16](#) PMID: [18282287](#)
98. Maccaferri M, Sanguineti MC, Corneti S, Ortega JLA, Salem MB, Bort J, et al. Quantitative trait loci for grain yield and adaptation of durum wheat (*Triticum durum* Desf.) across a wide range of water availability. *Genetics*. 2008; 178: 489–511. doi: [10.1534/genetics.107.077297](#) PMID: [18202390](#)
99. Stracke S, Haseneyer G, Veyrieras JB, Geiger HH, Sauer S, Graner A, et al. Association mapping reveals gene action and interactions in the determination of flowering time in barley. *Theor Appl Genet*. 2009; 118: 259–273. doi: [10.1007/s00122-008-0896-y](#) PMID: [18830577](#)
100. Breseghello F, Sorrells ME. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics*. 2006; 172: 1165–1177. PMID: [16079235](#)