

# Genome-wide Association Analysis of Kernel Morphology in Breeding Lines Derived from Synthetic Hexaploid Wheat in Qinghai Province, China

H. WANG<sup>1,2,4\*\*</sup>, D. CAO<sup>1,2,4\*\*</sup>, W. CHEN<sup>1,4\*</sup>, D. LIU<sup>3</sup>, B. LIU<sup>1,4</sup> and H. ZHANG<sup>1,2,4\*</sup>

<sup>1</sup>Qinghai Provincial Key Laboratory of Crop Molecular Breeding, Xining 810008, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup>Triticeae Research Institute, Sichuan Agricultural University, Chengdu, Sichuan 611130, China

<sup>4</sup>Key Laboratory of Adaptation and Evolution of Plateau Biota & Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining 810008

(Received 8 September 2017; Accepted 7 February 2018;

Communicated by A. Börner)

Wheat kernel morphology is a very important trait for wheat yield improvement. This is the first report of association analysis of kernel morphology traits in wheat breeding lines. In Qinghai, China, the research described here involved genome-wide association analysis in breeding lines derived from synthetic hexaploid wheat with a mixed linear model to identify the quantitative trait loci (QTLs) related to kernel morphology. The 8033 effective Diversity Array Technology (DArT) markers produced a genetic map of 5901.84 cM with an average density of 1.36 markers/cM. Population structure analysis classified 507 breeding lines into three groups by Bayesian structure analysis using unlinked markers. Linkage disequilibrium decay was observed with a map coverage of 2.78 cM. Marker-trait association analysis showed that 15 DArT markers for kernel morphology were detected, located on nine chromosomes, and explained 2.6%–4.0% of the phenotypic variation of kernel area (KA), kernel width (KW), kernel length (KL) and thousand-kernel weight (TKW). The marker 1139297 was related to both the KL and KA traits. Only six DArT markers were close to known QTLs. The parent SHW-L1 carried eight favored alleles, while other seven favored alleles were derived from elite common wheat cultivars. These QTLs, identified in elite breeding lines, should help us understand the kernel morphology trait better, and to provide germplasm for breeding new wheat cultivars for Qinghai Province or other regions.

**Keywords:** Genome-wide association, kernel morphology, Synthetic hexaploid wheat

## Introduction

Common wheat (*Triticum aestivum* L.,  $2n = 6x = 42$ , AABBDD) is the staple food for about 30% of the world's population. By 2050, global wheat production needs to have increased by a rate of 2% every year to meet demand, which is a huge challenge for wheat breeders (Hawkesford et al. 2013).

\*Corresponding authors; E-mail: cwj60905@163.com; hg Zhang@nwipb.cas.cn.

\*\*These authors contributed to this manuscript equally.

Wheat grain yield is underpinned by two numerical components, i.e. individual kernel weight and kernel number per m<sup>2</sup>. Over the past four decades, improvement of grain yield has come from increased number of kernels per m<sup>2</sup> or larger kernel sizes, due largely to the utilization of *Rht* genes in wheat breeding (Calderini et al. 2000). Since 1940, thousand-kernel weight (TKW) of Chinese wheat varieties has increased from an average of 31.5 g in the 1940s to 44.64 g in 2000, with an addition of 2.19 g every ten years over that period (Wang et al. 2012). TKW improvement is considered to be an important strategy for further improvement of yield potential in the Yellow and Huai Valleys in China and in Northwest Mexico (Xiao et al. 2012). High TKW can also improve seed germination rate, promote seedling growth and enhance the variety's ability to withstand natural stresses (Zhang et al. 2013). Furthermore, kernel size and shape in wheat significantly affect flour yield (Williams et al. 2013). Theoretical models predict that milling yield could be increased by optimizing kernel shape and size, with large spherical kernels being the optimum morphology (Evers et al. 1990).

TKW is a complex quantitative genetic trait. TKW varies under different natural environmental conditions, with different TKW-associated QTLs playing varied roles under different ecological conditions. The QTLs associated with TKW are distributed on almost all of the 21 different wheat chromosomes. The QTL profile should be determined under diverse ecological conditions so as to analyze the genetic regulatory network of this quantitative genetic trait more comprehensively. Qinghai Province is characterized by high altitude, thin cloud cover, long periods of sunshine, and perennial drought conditions, with little rain. Because of these high-stress ecological conditions, the yield of spring wheat in Qinghai Province ranges from 1.125 to 1.275 t/ha under normal crop management conditions, and the highest yield recorded (1.52 t/ha) was reported in 1978 (Chen 1993). Isolation of QTLs associated with kernel morphology in this region should help us to better understand the TKW trait.

To increase the genetic variation in common wheat, more and more synthetic hexaploid wheat (SHW) lines are being used in commercial wheat breeding, generating SHW-derived commercial cultivars since the first documented release of 'Chuanmai 42' in China, the latter being obtained by crossing a SHW from the International Maize and Wheat Improvement Center (CIMMYT) in Mexico with locally-adapted Chinese cultivars (Li et al. 2014). When crossed with locally-adapted wheat, some elite SHW-derived lines have been shown to yield up to 35% more than the best local checks (Dreccer et al. 2007; Ginkel and Ogbonnaya 2007; Yang et al. 2009; Ogbonnaya et al. 2013). The average yield of the SHW-derived cultivars was 13.5% higher than that of the non-synthetic cultivars (Tang et al. 2017). Recently, phenotypic and genetic diversity were evaluated, and favorable alleles or QTLs of yield-related traits were identified in the SHW varieties (lines) (Emebiri et al. 2015; Das et al. 2016; Zhang et al. 2017; Yan et al. 2017). In our lab, a synthetic hexaploid wheat SHW-L1 was used to insert alien alleles into common wheat. In this manuscript, 507 breeding lines from 17 hybrid combinations, using the same parent, SHW-L1, were used to identify the key QTLs of the kernel morphology traits in Qinghai Province, to identify the elite alleles from the SHW, and to develop strategies to improve these advanced lines.

## Materials and Methods

### *Plant material*

A total of 507 F<sub>10</sub> elite lines were used in this study (Table S1\*). The lines were derived from 17 hybrid combinations and shared the same parental synthetic hexaploid wheat, SHW-L1, which was synthesized from a cross between the *Triticum turgidum* L. (2n = 4x = 28, AABB) line AS2255 with *Aegilops tauschii* Coss. (2n = 2x = 14, DD) accession AS60 (Zhang et al. 2004). These lines were developed by a top-cross of SHW-L1 with common wheat cultivars. Artificial selection was carried out at every generation to obtain plants with desirable agronomic traits, according to the selection requirements of local commercial cultivars. All the 507 SHW-L1-derived lines exhibited a yield similar to or higher than that of local commercial cultivars in Xining, in Qinghai Province.

### *Phenotyping kernel morphology*

These breeding lines were planted in a randomized complete block design with three replicates, in 2013 and 2014 in Xining (101°74' E, 36°56' N), Qinghai, China. Each replicate row was 2 m long, with 20 cm row-to-row distance. Twenty representative plants were selected from each replicate for measuring kernel length (KL), kernel width (KW), kernel ratio (KL/KW), kernel area (KA) and factor from density (FFD) after the plants were harvested. All lengths were measured directly in a high flux Marvin seed analyzer (GTA Sensorik GmbH, Neubrandenburg, Germany). TKW was determined by weighing triplicate batches of 1000 kernels each. Data analyses were performed using statistical software SPSS version 17.0 (SPSS, Chicago, IL, USA).

### *DNA extraction and DArT genotyping*

Leaf tissue for DNA extraction was collected and stored at -80 °C. Genomic DNA was extracted using the cetyltrimethylammonium bromide method as outlined by Yan et al. (2002) <sup>®</sup>. A wheat Diversity Array Technology (DArT) sequence array consisting of 39,188 random sequences (<http://www.triticarte.com.au>) was used in DArT genotyping. Procedures for hybridization of genomic DNA to the DArT-sequence array, image analysis and polymorphism scoring were as described by Akbari et al. (2006). The information of DArT markers was obtained from Triticarte Pty. Ltd. (Canberra, Australia; <http://www.triticarte.com.au>). The markers were filtered according to the conditions (reproducibility >95%, chromosomal location information, minor allele frequency >5%, percentage of missing data <10%) for population structure, linkage disequilibrium, and association analysis.

\*Further details about the Electronic Supplementary Material (ESM) can be found at the end of the article.

### *Population structure, linkage disequilibrium, and association analysis*

Population structure analysis was estimated with STRUCTURE v 2.2 software (Pritchard et al. 2000), based on the genotyping data. Linkage disequilibrium (LD) among markers was measured with TASSEL 3.0 software (Bradbury et al. 2007). The methodology and data analysis, concerning population structure and LD, were as described by Neumann et al. (2011). Association analysis was calculated with the mixed linear model (MLM) with the TASSEL 3.0 software (Bradbury et al. 2007). The MLM took Q-matrix and kinship-matrix into account to reduce both Type I and Type II errors (Yu et al. 2006). The Q-matrix was derived from STRUCTURE v 2.2 (Pritchard et al. 2000) and the kinship-matrix was calculated using TASSEL 3.0 software. Different agronomic traits have different significance criteria, 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). The map of each group was drawn using MapChart v.2.2 (<http://www.biometris.nl/uk/Software/MapChart/>).

## Results

### *Phenotypic characters*

KL, KW, KL/KW, FFD and TKW were measured to evaluate kernel morphology. The maximum values of KL, KW, KL/KW, FFD and TKW were 8.3 mm, 4.0 mm, 3.0, 24.6 mm<sup>2</sup>, 0.003028 and 58.8 g, respectively, while the minimum values were 5.8 mm, 2.3 mm, 1.6, 13.6 mm<sup>2</sup>, 0.000494 and 28.1 g, respectively (Table 1). The maximum values of KW, KA and TKW were almost twice those of the corresponding minimum values (Table 1). All skewness and kurtosis values for KL, KW, KA and TKW were less than 1 (Figure S1, Table 1). The kurtosis value of KL/KA was greater than one, while both the skewness and kurtosis values of FFD were greater than one. As a result, KL, KW, KA and TKW could be analyzed with marker-trait association (MTA) analysis, while KL/KA and FFD could not. The variability of kernel morphology within the lines are shown in Table S2. Pearson's coefficient of correlation was calculated to determine the relationships among the agronomic traits (Table S3). The only significant positive correlations existed among the four traits KL, KW, KA and TKW. The highest positive correlation ( $r = 0.850$ )

Table 1. The kernel morphology of the breeding lines

Traits	Average±SD	Min	Max	Coefficient of Variation CV %	Heritability H <sup>2</sup>	Skewness	Kurtosis
KL	6.76±0.33	5.8	8.3	4.88	0.92	-0.104	0.556
KW	3.40±0.26	2.3	4.0	7.65	0.68	-0.093	0.250
KL/KW	2.00±0.16	1.6	3.0	8.0	0.89	-0.93	1.640
KA	19.35±1.84	13.6	24.6	9.51	0.62	-0.108	0.603
FFD	0.0018±0.00028	0.000494	0.003028	15.56	0.90	1.461	1.178
TKW	44.03±8.7	28.1	58.8	19.64	0.72	-0.066	0.433

was observed between KA and TKW, followed by  $r = 0.826$  between KA and KL (Table S3). The variables KL/KW and FFD exhibited significant negative correlations with the four remaining traits. The highest negative correlation ( $-0.774$ ) occurred between KW and KL/KW (Table S3).

### Marker coverage and polymorphism

A total of 39,188 DArT markers were used to genotype the 507 breeding lines. A total of 24,872 markers had reproducibility values  $>95\%$ . Taking into account information such as the chromosome location, the minor allele frequency ( $>5\%$ ), and the percentage of missing data ( $<10\%$ ), 8033 markers remained for further population structure, LD and association analysis (Table S4). These DArT markers, integrated into the framework genetic map, covered a total genetic distance of 5901.84 cM over the entire 21 chromosomes, with an average density of 1.36 markers/cM (Table S5). The average density of markers for B-genome chromosomes was the highest (2.05 markers/cM), while the lowest was for D-genome chromosomes (0.61 markers/cM) (Table S5). Polymorphic information content (PIC) values ranged from 0.096 to 0.497, with an average of 0.36.

### Population structure

396 DArT markers (one marker was chosen every 10 cM) were selected from all chromosomes for population structure analysis. When  $K = 3$ , the  $\Delta K$  value was highest (3577.078)

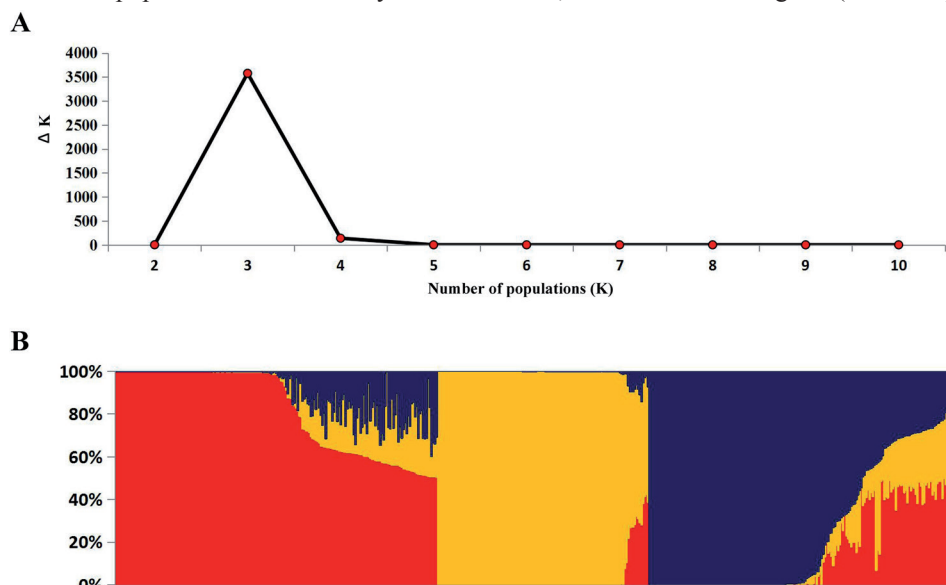


Figure 1. Population structure of synthetic hexaploids based on DArT markers. a) Plot of the average logarithm of the probability of data  $\Delta K$  ( $\Delta K$ ), as second rate of change of the number of assumed subgroups (K), with K allowed to range from 2 to 10. B) Membership co-efficient (Q value) where each horizontal line represents one wheat line, and partitioned into three sub-populations

(Fig. 1A). Therefore, a K value of 3 was regarded to be the most suitable number of subpopulations. The 507 breeding lines were grouped into three subpopulations (Fig. 1B). Subpopulation 1 was the largest group and comprised 240 breeding lines, which were from 13 combinations (1, 3, 4, 6, 8, 9, 10, 11, 12, 14, 15, 16 and 17). Subpopulation 2 consisted of 131 breeding lines and was mainly composed of lines from combinations 5 and 7. Subpopulation 3 included 136 lines and was from combinations 2 and 13. Breeding lines from the same combination were allocated to the same subpopulation.

### LD analysis

LD was estimated by  $r^2$  at  $P \leq 0.001$  for all pairs of the 8033 DArT markers. At a genome-wide level, almost 77.6% of all pairs of loci were in significant LD (Table S6). The average  $r^2$  of genome-wide LD was 0.27. DArT markers, assigned to their map position, were further used to estimate inter- and intra-chromosomal LD. About 13,944 (3.48%) inter-chromosomal pairs of loci were in significant LD ( $r^2 \leq 0.001$ ), with an average  $r^2$  of 0.12, while 29,655 (78.1%) intra-chromosomal pairs of loci were in significant LD with an average  $r^2$  of 0.28. The extent and distribution of LD were graphically displayed by plotting intra-chromosomal  $r^2$  values for loci in significant LD at  $P \leq 0.001$  against the genetic distance in cM, and a second-degree locally weighted polynomial regression-based (LOESS) curve was fitted (Fig. 2, red line). The LD critical threshold ( $r^2 = 0.34$ ), estimated from LD distribution of pairs of intra-chromosomal DArT markers, is indicated by the dashed horizontal line in Fig. 2 (Rasheed et al. 2014). Thus, the map coverage of 2.78 cM was deemed appropriate to perform a genome-wide association analysis on the population.

### MTAs

MTAs in 507 breeding lines were identified using the mixed linear model (MLM), and a total of 15 DArT markers were detected for agronomic traits, and explained 2.6–4.0% of the kernel morphology variance (Fig. 3; Fig. S2; Table 2). The number of MTAs for KL,

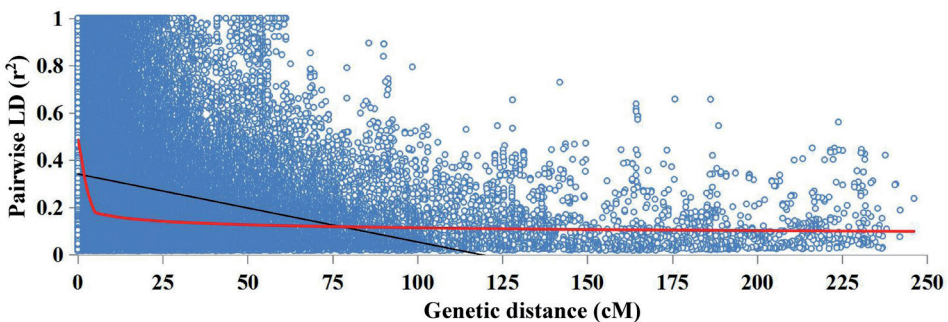


Figure 2. Scatterplot of the LD statistic  $r^2$  as a function of genetic distance (cM) between pairs of DArT markers in SHWs. The locally weighted polynomial regression-based (LOESS) (red line) representing decay of  $r^2$  against long genetic distance is illustrated for each genome. The baseline (black line) intersection with the LOESS curve was considered as the estimate of the extent of LD (Rasheed et al. 2014). LD critical threshold estimated from LD distribution of pairs of unlinked DArT markers is indicated by the solid horizontal line

Table 2. Marker-trait associations (MTAs) for kernel morphology

Trait	Marker	Chromosome	Position	P	R <sup>2</sup> %	Reported in Reference	Favored allele	The alleles in SHW-L1
KL	1110840	1B	287.9	3.06E-04	2.6		1	0
	1139297	5A	8.7	9.99E-05	3.3		0	0
	1724620	6B	120.0	2.32E-04	2.8		0	1
KW	998369	1D	152.7	2.19E-04	2.9		0	1
	1665829	3D	9.5	3.88E-05	3.5		1	0
	1107696	4B	66.6	1.68E-04	2.8	<i>QKW-4B</i> (Cui et al. 2015)	1	1
	1099650	5D	52.6	5.18E-05	3.4	<i>QKwpp-5D</i> (Cui et al. 2014)	0	0
	1076471	6A	25.1	4.33E-04	2.7	<i>QYLD-6A</i> (Bordes et al. 2014)	0	1
KA	3023348	6A	33.2	3.73E-04	2.9		0	1
	1122839	6D	102.8	2.72E-05	4.0		0	0
	1204996	6D	111.5	1.15E-04	3.1		1	1
	1139297	5A	8.7	8.24E-05	3.4		0	0
TKW	1100949	3D	161.4	1.42E-04	3.2	<i>QTKw-3D</i> (Rasheed et al. 2014)	1	0
	1210312	6A	1.2	2.72E-04	2.7	<i>QYld-6A</i> (Bordes et al. 2014)	0	0
	1074463	6A	177.2	1.15E-04	3.0	<i>QGwe.ipk-6A</i> (Börner et al. 2002)	0	0

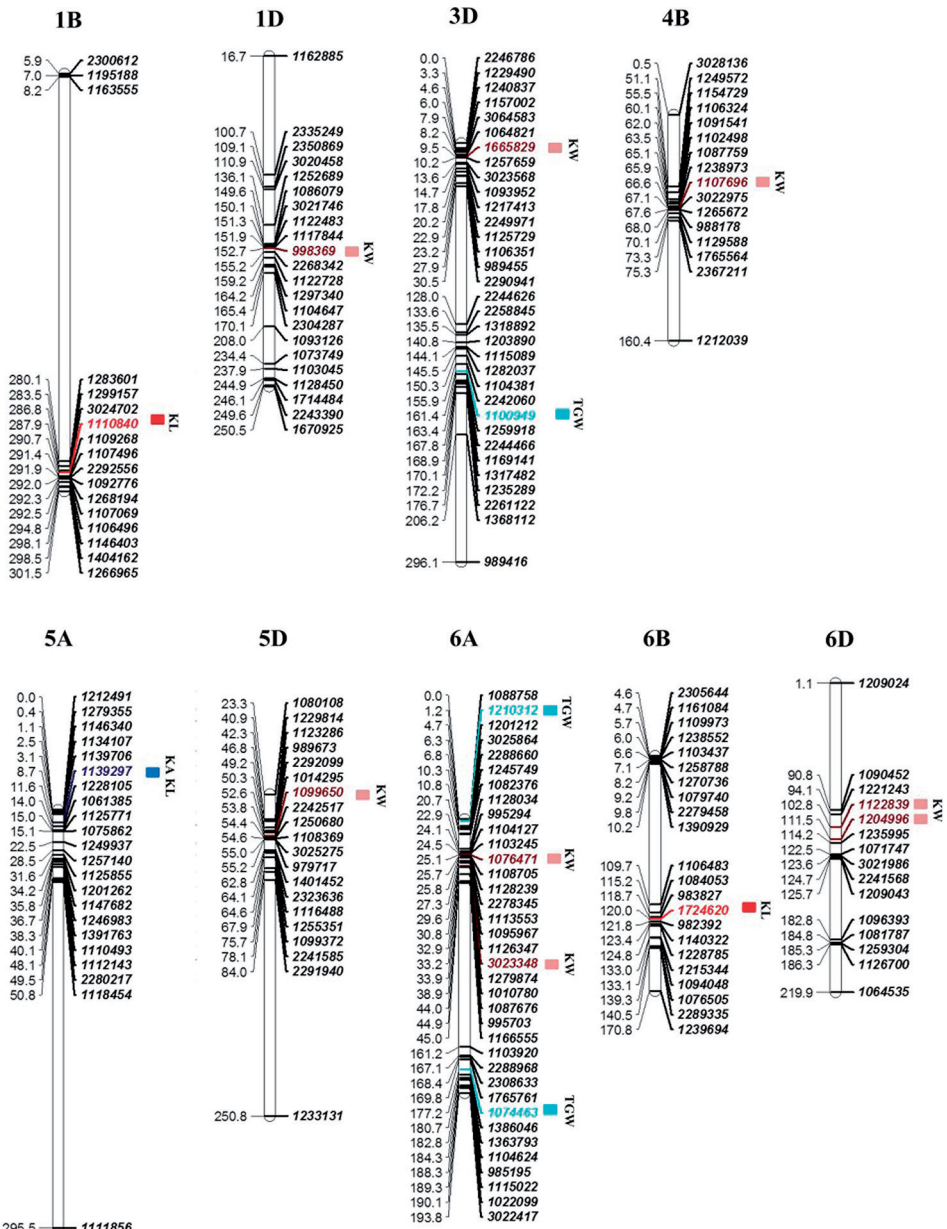


Figure 3. DArT consensus linkage map of chromosomes showing marker-traits associations for grain morphology in the breeding lines population. MTAs are projected as different color solid bars



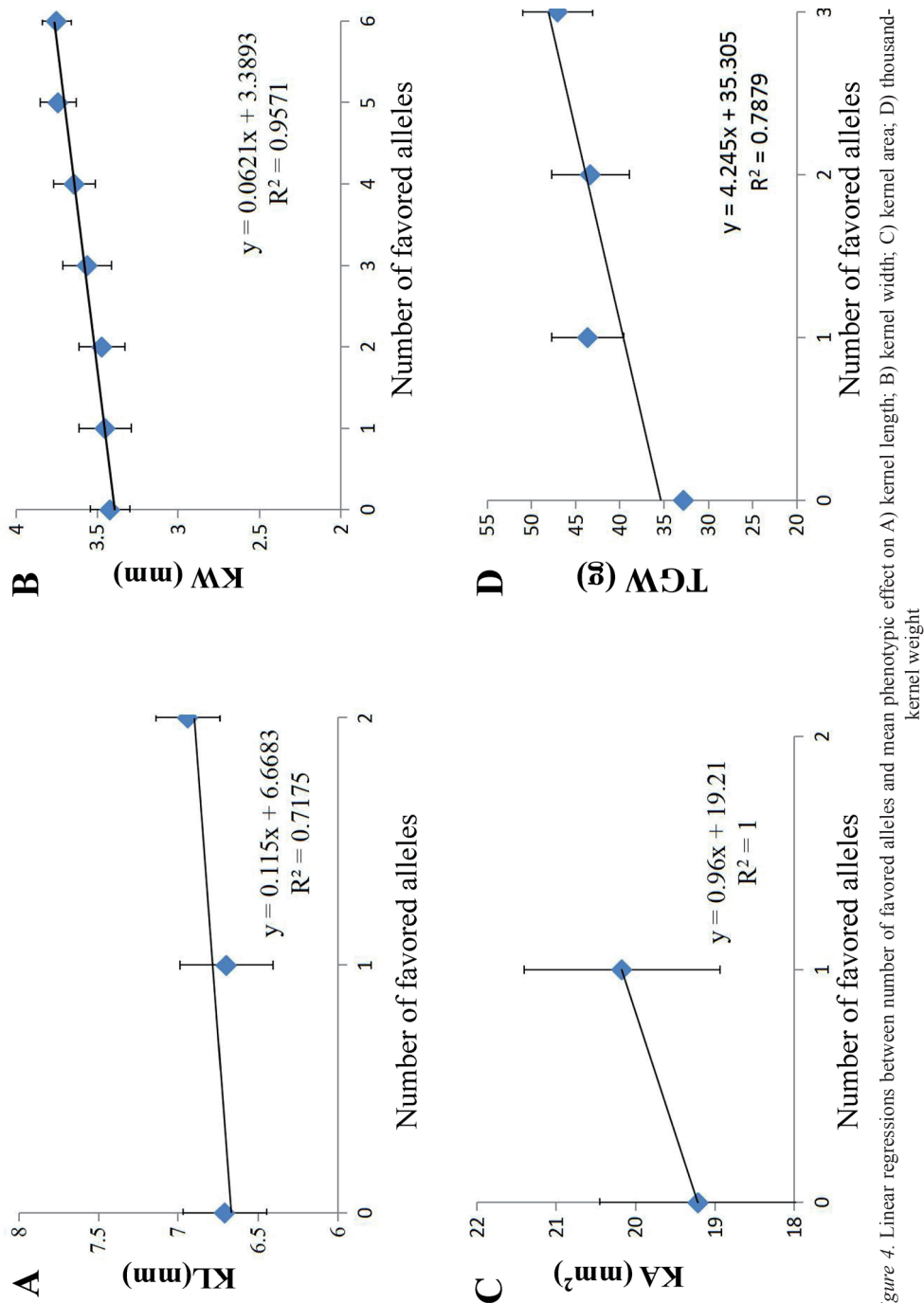


Figure 4. Linear regressions between number of favored alleles and mean phenotypic effect on A) kernel length; B) kernel width; C) kernel area; D) thousand kernel weight

KW, KA and TKW were 3, 8, 1 and 3, respectively. The MTAs were located on nine chromosomes (1B, 1D, 3D, 4B, 5A, 5D, 6A, 6B and 6D). Eight MTAs for KW were identified on six of the 21 chromosomes and explained 2.7%–4.0% of the phenotype variation; one MTA for KW, located on chromosome 6D (102.8 cM), was highly significant ( $P = 2.72 \times 10^{-5}$ ), and explained 4.0% of the variation. Only one MTA was identified for the KA trait. This MTA resided on chromosome 5A, and explained 3.4% of the variation in KA. The MTA was also identified in the KL trait, and controlled 3.3% of the KL trait variation. The two remaining MTAs for KL were found on chromosomes 1B and 6B, and explained 2.6% and 2.8%, respectively of the phenotype variation. The three MTAs for TKW explained 3.2%, 2.7% and 3.0% of the phenotypic variation and were localized on chromosomes 3D and 6A.

#### *Compared with the reported MTAs for kernel morphology*

The MTAs for KL and KA had no QTLs reported nearby, while all the MTAs for TKW had been reported before (Table 2) (Rasheed et al. 2014; Bordes et al. 2014; Börner et al. 2002). Three MTAs for KW traits were close to known QTLs (Table 2) (Cui et al. 2014, 2015; Bordes et al. 2014), while the remaining four MTAs had not been reported previously. For investigating the origin of the elite alleles, the genotype of SHW-L1 was compared with the favored alleles. SHW-L1 carried eight favored alleles including 1139297, 1724620, 998369, 1665829, 3023348, 1122839, and 1204996 (Table 2). The remaining six favored alleles were derived from the elite common wheat cultivars.

#### *Relationship between kernel phenotype and the number of favored alleles*

A linear relationship was observed between the copy number of favored alleles and each of the four kernel traits, KL, KW, KA and TKW, where the addition of every favorable allele in a variety additively contributed to enhance the phenotypic trait (Fig. 4). The linear correlation coefficient for KW ( $P < 0.05$ ) was the highest. Breeding with fewer alleles should reduce the linear correlation coefficients. Some breeding lines carried all the favored alleles for KA and TKW, while no breeding lines carried all the favored alleles for KL and KW. Most lines carried two favored alleles for KL, while most lines carried six of the eight favored alleles for KW.

### **Discussion**

MTA analysis, to identify the QTLs responsible for agronomic traits, is usually carried out on natural populations. In this study, breeding lines with stable agronomic traits suited to local conditions were used to identify the QTLs present in this population. In total, 14 MTAs for KL, KW, KA, and TKW were identified in 507 breeding lines. Previously, some QTLs or MTAs for KL, KW, KA, and TKW were identified (Börner et al. 2002; Bordes et al. 2014; Rasheed et al. 2014; Cui et al. 2014, 2015; Yan et al. 2017). The MTAs for KL and KA had no QTLs reported nearby, while all the MTAs for TKW (1100949,

1210312, 1074463) had been reported before (Rasheed et al. 2014; Bordes et al. 2014; Börner et al. 2002). Three MTAs (1107696, 1099650, 1076471) for KW traits were close to known QTLs (Cui et al. 2014, 2015; Bordes et al. 2014), while the remaining four MTAs had not been reported previously.

These QTLs could be used for further improvement of the breeding lines, and for breeding new wheat cultivars in this and other regions. These breeding lines, from different hybrid combinations, could be regarded as a natural population, but the QTLs could not be analyzed in the lines derived from the same combination because the population structure was not suitable for association analysis (data not shown). Few QTLs identified in our manuscript have been reported before because of the nature of the population and the particular experimental conditions.

The LD is influenced by recombination rate, allele frequency, population structure and selection pressure (Flint-Garcia et al. 2003). Rasheed et al. (2014), using 231 SHWs with 834 DArT markers, revealed that the LD decay distance was 9cM. In another study, of 91 SHWs with 395 DArT markers, Emebiri et al. (2010) showed that the LD decay distance was up to 15 cM. In our study, the LD decay distance was about 2.78 cM for the whole genome. This comparison shows that the LD decay distance in the current study was short. This indicated that extensive recombination of A, B and D genomes had occurred in the top-cross procedure, and that linkage was broken, presumably due to hybridization and selection or to the high marker density in our study.

TKW is a complex agronomic trait, determining wheat production, while KL, KW and KA are the biological traits, exhibiting positive relationships with TKW, KL, KW and KA should be controlled by the genes directly, while TKW should be controlled indirectly. In this study, several MTAs were identified, but the MTAs were specific to individual traits, except for one MTA, 1139297, which controlled both KL and KA. The MTAs of the biological traits KL, KW and KA could not be detected in TKW marker-trait associations because of the confounding effect. Though the MTAs could not be detected, MTA accumulation in the wheat lines will improve the biological traits KL, KW and KA, and influence the TKW value eventually. Unfortunately, no breeding lines carried all of the favored alleles, which suggests that some favored alleles only existed in certain combinations. Meanwhile, these imperfect breeding lines should give us the chance to develop improved breeding lines by accumulating more of the favored alleles. The next step should be to assemble all the MTAs into a single breeding line, so that new improved cultivars can be bred, based on the genotyping of these breeding lines.

### Acknowledgements

This research was financially supported by the West Light Foundation of the Chinese Academy of Sciences, the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA08030106), the Ministry of Science and Technology of China (2016YFD0100500) and the Science and Technology Planning Project of Qinghai Province, China (2013-Z-942Q and 2017-ZJ-Y14).

## References

- Akbari, M., Wenzl, P., Caig, V., Carling, J., Xia, L., Yang, S., Uszynski, G., Mohler, V., Lehmensiek, A., Kuchel, H. 2006. Diversity arrays technology (DARt) for high-throughput profiling of the hexaploid wheat genome. *Theor. Appl. Genet.* **113**:1409–1420.
- Bordes, J., Goudemand, E., Duchalais, L., Chevarin, L., Oury, X.F., Heumez, E., Lapierre, A., Perretant, M.R., Rolland, B., Beghin, D., Laurent, V., Gouis, J.L., Storlie, E., Robert, O., Charmet, G. 2014. Genome-wide association mapping of three important traits using bread wheat elite breeding populations. *Mol. Breed.* **4**:755–768.
- Börner, A.E., Schumann, A., Fürste, H., Cöster, B., Leithold, M.S., Röder, W.E. 2002. Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* **105**:921–936.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics.* **23**:2633–2635.
- Calderini, D.F., Reynolds, M.P., Calderini, D.F., Reynolds, M.P. 2000. Changes in grain weight as a consequence of de-graining treatments at pre- and post-anthesis in synthetic hexaploid wheats. *Aust. J. Plant Physiol.* **27**:183–191.
- Chen, D.Z. 1993. New progress of wheat cultivation in China. Agriculture Press, Beijing, pp. 327–340 (in Chinese).
- Cui, F., Fan, X.L., Chen, M., Zhang, N., Zhao, C.H., Zhang, W., Han, J., Ji, J., Zhao, X., Yang, L., Wang, T., Li, J.M. 2015. QTL detection for wheat kernel size and quality and the responses of these traits to low nitrogen stress. *Theor. Appl. Genet.* **129**:469.
- Cui, F., Zhao, C.H., Ding, A.M., Li, J., Wang, L., Li, X.F., Bao, Y.G., Li, J.M., Wang, H.G. 2014. Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations. *Theor. Appl. Genet.* **127**:659–675.
- Das, M.K., Bai, G., Mujeeb-Kazi, A., Rajaram, S. 2016. Genetic diversity among synthetic hexaploid wheat accessions (*Triticum aestivum*) with resistance to several fungal diseases. *Genet. Resour. Crop. Ev.* **63**:1285–1296.
- Dreccer, M.F., Borgognone, M.G., Ogonnaya, F.C., Trethowan, R.M., Winter, B. 2007. CIMMYT-selected derived synthetic bread wheats for rainfed environments: Yield evaluation in Mexico and Australia. *Field Crops Res.* **100**:218–228.
- Emebiri, L.C., Ogonnaya, F.C. 2015. Exploring the synthetic hexaploid wheat for novel sources of tolerance to excess boron. *Mol. Breed.* **35**:1–10.
- Emebiri, L.C., Oliver, J.R., Mrva, K., Mares, D. 2010. Association mapping of late maturity  $\alpha$ -amylase (LMA) activity in a collection of synthetic hexaploid wheat. *Mol. Breed.* **26**:39–49.
- Evers, A.D., Cox, R.I., Shaheedullah, M.Z., Withey, R.P. 1990. Predicting milling extraction rate by image analysis of wheat grains. *Aspects Appl. Biol.* **25**:417–426.
- Flint-Garcia, S.A., Thornsberry, J.M., Buckler, E.S. 2003. Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Physiol.* **54**:357–374.
- Ginkel, M.V., Ogonnaya, F. 2007. Novel genetic diversity from synthetic wheats in breeding cultivars for changing production conditions. *Field Crops Res.* **104**:86–94.
- Hawkesford, M.J., Araus, J.L., Park, R., Calderini, D., Miralles, D., Shen, T., Zhang, J., Parry, M.A.J. 2013. Prospects of doubling global wheat yields. *Food and Energy Security.* **2**:34–48.
- Li, J., Wan, H., Yang, W. 2014. Synthetic hexaploid wheat enhances variation and adaptive evolution of bread wheat in breeding processes. *J. Syst. Evol.* **52**:735–742.
- Neumann, K., Kobiljski, B., Denčić, S., Varshney, R., Börner, A. 2011. Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Mol. Breed.* **27**:37–58.
- Ogonnaya, F.C., Abdalla, O., Mujeeb-Kazi, A., Kazi, A.G., Xu, S.S., Gosman, N., Lagudah, E.S., Bonnett, D., Sorrells, M.E., Tsujimoto, H. 2013. Synthetic hexaploids: harnessing species of the primary gene pool for wheat improvement. John Wiley & Sons, Inc. pp. 35–122.
- Pritchard, J.K., Stephens, M., Rosenberg, N.A., Donnelly, P. 2000. Association mapping in structured populations. *The American Journal of Human Genetics.* **67**:170–181.

- Rasheed, A., Xia, X.C., Ogonnaya, F., Mahmood, T., Zhang, Z.W., Mujeeb-Kazi, A. He, Z.H. 2014. Genome-wide association for grain morphology in synthetic hexaploid wheats using digital imaging analysis. *BMC Plant Biology*. **14**:128.
- Sukumaran, S., Xiang, W., Bean, S.R., Pedersen, J.F., Kresovich, S., Tuinstra, M.R., Tesso, T.T., Hamblin, M.T., Yu, J.M. 2012. Association mapping for grain quality in a diverse sorghum collection. *Plant Genome*. **5**:126–135.
- Tang, Y., Wu, X.L., Li, C.S., Yang, W.Y., Huang, M.B., Ma, X.L., Li, S.Z. 2017. Yield, growth, canopy traits and photosynthesis in high-yielding, synthetic hexaploid-derived wheats cultivars compared with non-synthetic wheats. *Crop & Pasture Science* **68**:115–125.
- Wang, L.F., Ge, H.M., Hao, C.Y., Dong, Y.S., Zhang, X.Y. 2012. Identifying Loci Influencing 1,000-Kernel Weight in Wheat by Microsatellite Screening for Evidence of Selection during Breeding. *PLoS One*. **7**: e29432.
- Williams, K., Munkvold, J., Sorrells, M. 2013. Comparison of digital image analysis using elliptic Fourier descriptors and major dimensions to phenotype seed shape in hexaploid wheat (*Triticum aestivum* L.). *Euphytica*. **190**:99–116.
- Xiao, Y.G., Qiang, Z.G., Wu, K., Liu, J.J., Xia, X.C., Ji, W.Q., He, Z.H. 2012. Genetic gains in grain yield and physiological traits of winter wheat in Shandong province, China, from 1969 to 2006. *Crop Sci*. **52**:44–56.
- Yang, W., Liu, D., Li, J., Zhang, L., Wei, H., Hu, X., Zheng, Y., He, Z., Zou, Y. 2009. Synthetic hexaploid wheat and its utilization for wheat genetic improvement in China. *J. Genet. Genomics*. **36**:539–546.
- Yan, L., Liang, F., Xu, H., Zhang, X., Zhai, H., Sun, Q., Ni, Z. 2017. Identification of QTL for grain size and shape on the D genome of natural and synthetic allohexaploid wheat with near-identical AABB genomes. *Front. Plant Sci*. **8**:1705.
- Yan, Z., Wan, Y., Liu, K., Zheng, Y., Wang, D. 2002. Identification of a novel HMW glutenin subunit and comparison of its amino acid sequence with those of homologous subunits. *Chinese Sci. Bull.* **47**:220–225.
- Yu, J., Pressoir, G., Briggs, W.H., Vroh, B.I., Yamasaki, M., Doebley, J.F., McMullen, M.D., Gaut, B.S., Nielsen, D.M., Holland, J.B., Kresovich, S., Buckler, E.S. 2006. A unified mixed-model 24 method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* **38**:203–208.
- Zhang, H.X., Zhang, F.N., Li, G.D., Zhang, S.N., Zhang, Z.G., Ma, L.J. 2017. Genetic diversity and association mapping of agronomic yield traits in eighty six synthetic hexaploid wheat. *Euphytica*, **213**:111.
- Zhang, L., Liu, D., Yan, Z., Lan, X., Zheng, Y., Zhou, Y. 2004. Rapid changes of microsatellite flanking sequence in the allopolyploidization of new synthesized hexaploid wheat. *Science in China*. **47**:553–561.
- Zhang, K.P., Wang, J.J., Zhang, L.Y., Rong, C.W., Zhao, F.W., Peng, T., Li, H.M., Cheng, D.M., Liu, X., Qin, H.J. 2013. Association Analysis of Genomic Loci Important for Grain Weight Control in Elite Common Wheat Varieties Cultivated with Variable Water and Fertiliser Supply. *PLoS One*. **8**:e57853–e57853

### Electronic Supplementary Material (ESM)

Electronic Supplementary Material (ESM) associated with this article can be found at the website of CRC at <https://akademai.com/loi/0806>

Electronic Supplementary *Table S1*. The pedigree of the 507 elite lines

Electronic Supplementary *Table S2*. The grain morphology of the breeding lines in 15 groups

Electronic Supplementary *Table S3*. Pearson's co-efficient of correlation for important agronomic traits in wheat lines

Electronic Supplementary *Table S4*. The information of DArT markers

Electronic Supplementary *Table S5*. The distribution of DArT markers on wheat chromosomes

Electronic Supplementary *Table S6*. An overview of LD among whole panel of lines

Electronic Supplementary *Figure S1*. The frequency distribution of phenotype traits in 507 lines. A) Kernel length B) Kernel width C) Kernel area D) Thousand kernel weight