Mol Breeding (2018) 38: 146
https://doi.org/10.1007/s11032-018-0906-y

A genome-wide associate study reveals favorable alleles conferring apical and basal spikelet fertility in wheat (*Triticum aestivum* L.)

Weiping Shi · Linqi Yue · Jingye Cheng · Jiahui Guo · Lulu Li · Kaidi Xie · Jiarui Dong · Yanhao Xu · Jie Guo · Meixue Zhou

Received: 25 July 2018 / Accepted: 14 November 2018 / Published online: 20 November 2018

Abstract Kernel number per spike (KNPS) is one of the key factors affecting wheat yield, which can be significantly reduced by lower fertility or sterility of the apical and basal spikelets. In this study, the spikelet number per spike (SNPS), thousand kernel weight (TKW), KNPS, total grain numbers of the top three apical spikelets (GNAS), and total grain numbers of the bottom three basal spikelets (GNBS) of 212 wheat lines were recorded from five different environmental conditions. These 212 accessions were genotyped using the 9K iSelect SNP Beadchip. A total of 3269 SNP markers were used for genome-wide association analysis (GWAS). One hundred twelve significant marker-trait associations (MTAs) were identified. Twenty-two MTAs were identified in at least two environments and two of them showed association with two or more grain setting properties. Different loci showed an additive effect with both GNAS and GNBS being much higher in the lines with more favorite alleles. Two SNP loci, *wsnp_Ex_c31799_40545376* and *wsnp_BF293620A_Ta_2_3*, showed the largest effects on increasing KNPS through improved fertility of apical and basal spikelets, respectively. These MTAs have the potential to be used in future marker-assisted selection.

Keywords Wheat · Spikelet fertility · GWAS

Weiping Shi, Linqi Yue and Jingye Cheng contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11032-018-0906-y) contains supplementary material, which is available to authorized users.

W. Shi · L. Yue · J. Cheng · J. Guo · L. Li · K. Xie · J. Dong · J. Guo (✉) · M. Zhou (✉)
College of Agronomy, Shanxi Agricultural University, Taigu 030801, China
e-mail: nxgj1115326@163.com
e-mail: Meixue.Zhou@utas.edu.au

J. Cheng
College of Agronomy, Yangzhou University, Yangzhou 225009, China

Y. Xu · M. Zhou
Hubei Collaborative Innovation Centre for Grain Industry, Yangtze University, Jingzhou 434025, China

M. Zhou
School of Land and Food, University of Tasmania, Hobart 7250, Australia

Introduction

Wheat is the most important food crop and ranks first in harvested area, total production, and traded volume worldwide (Hawkesford et al. 2013; Tripathi et al. 2016). Wheat production largely contributes to food security, socioeconomic development, and living standards (Piao et al. 2010). Kernel number per spike (KNPS) is one of the key factors affecting yield (Fischer 2008; Reynolds et al. 2009; Gao et al. 2017). An unbalanced distribution of grains per spikelet along the spike (top, center, and bottom of a spike) has been widely reported and the fertility of the apical and basal spikelets showed greater effects on KNPS than the middle spikelets in

wheat spikes (Ferrante et al. 2013a; Guo and Schnurbusch 2015). Therefore, breeding wheat with higher fertility of the apical and basal spikelets could increase KNPS and thus yield (Acreche et al. 2008; Zheng et al. 2016).

Lower fertility or sterility of apical and basal spikelets is commonly observed in cereal crops such as wheat and rice (Satoh-Nagasawa et al. 2006; Meng et al. 2007; Gallavotti et al. 2011; Guo and Schnurbusch 2015). Genetic variations have been found in different genotypes with most varieties showing a lower fertility of the apical and basal spikelets. To understand the mechanism of this phenomenon, many morphological and physiological studies of spike development have been conducted (Bancal 2009; Shitsukawa et al. 2009; González et al. 2011; Ferrante et al. 2013b; González-Navarro et al. 2015). Langer and Hanif (1973) found unsynchronized development of spikelets dependent on the position within the spike, where the two basal spikelets developed much slower leading to lower fertility or complete sterility of basal spikelets. It was also found that spikelet fertility was largely affected by environmental factors such as planting density (Mishra and Mohapatra 1987), trace element (Rerkasem and Jamjod 1997), sowing time (Saifuzzaman et al. 2008), and drought (Dencic et al. 2000). A good crop management is essential to maximize the number of fertile florets and improve grain set numbers in apical and basal spikelets (Ferrante et al. 2010, 2012; Dreccer et al. 2014; Zheng et al. 2014).

Many studies have been conducted in cereal crops to identify genes or quantitative trait loci (QTL) controlling the fertility of apical and basal spikelets (Yamagishi et al. 2004; Li et al. 2009; Tan et al. 2011; Cheng et al. 2011; Akter et al. 2014). In rice, Yamagishi et al. (2004) located three QTL affecting pre-flowering basal floret abortion on chromosomes 1, 10, and 11, respectively. A candidate gene *Short panicle1* was isolated and mutation of this gene could cause significant reductions in basal floret numbers (Li et al. 2009). An interactive effect was also found between different QTL for apical spikelet fertility (Tan et al. 2011). *qPAA8*, a gene controlling panicle development in rice, has been fine mapped to the 68 kb zone on chromosome 8 (Cheng et al. 2011). There are limited studies in wheat on the inheritance of grain set in apical and basal spikelets (Guo et al. 2015; Guo et al. 2017). After performing genome-wide association studies (GWAS) of 16 floret fertility traits in 210 European winter wheat accessions, Guo et al. (2017) proposed a genetic network underlying floret fertility and related traits, nominating determinants for improved yield performance. A full understanding of the genetics of grain setting at the molecular level is needed for breeders to improve apical and basal floret fertility.

The aim of this study was to identify MTAs for the fertility of three apical and basal floret. The grain numbers in apical and basal spikelets of 212 wheat varieties were collected from five environmental conditions (years/locations). Genome-wide association studies revealed several MTAs controlling grain numbers in apical and basal spikelets. These MTAs have the potential to be used in future fine mapping, cloning, and marker-assisted selection.

Materials and methods

Plant materials

The materials consisted of 212 wheat varieties, including 200 from China, 3 from Italy, 1 from Japan, 1 from Pakistan, and 7 with unknown origins (Table S1). The Chinese varieties were from Jiangsu (63), Henan (22), Shaanxi (21), Shandong (19), Sichuan (15), Beijing (12), Anhui (11), Hunan (9), Hebei (7), Hubei (6), Shanxi (4), Fujian (4), Gansu (3), Guizhou (2), Jiangxi (1), and Zhejiang (1).

Phenotyping

All genotypes were planted at three locations (Jingzhou (JZ) in Hubei province; Yangzhou (YZ) in Jiangsu; and Xinxiang (XX) in Henan) in three growing seasons (2013–2014 (14), 2014–2015 (15), and 2015–2016 (16)). The environments were designated as 14JZ, 14YZ, 15JZ, 15YZ, and 16XX, respectively. Field experiments used randomized block designs with three replications. Each line was planted in five 2-m-long rows with a row spacing of 0.2 m. Forty seeds were planted in each row that were thinned back to 30 per row after germination giving a final plant density of 75 plants/m^2. Field management followed local practices. Seedling numbers were thinned to about 30 per row at early seedling stage. The traits recorded included spikelet number per spike (SNPS), thousand kernel weight (TKW), KNPS, grain numbers of the top three apical spikelets (GNAS), and grain numbers of the bottom

Table 1 Descriptive statistics for nine phenotypic traits assessed in five environments

Trait	14JZ					14YZ					15JZ				
	Mean ± SD	Min	Max	CV (%)	h^2	Mean ± SD	Min	Max	CV (%)	h^2	Mean ± SD	Min	Max	CV (%)	h^2
SNPS	21.32 ± 1.39	17.20	25.50	6.50	52.76	21.59 ± 1.43	16.85	25.10	6.62	50.52	20.47 ± 1.31	15.85	23.50	6.42	50.74
TKW	39.19 ± 5.14	27.50	53.95	13.11	49.13	38.06 ± 5.35	27.05	51.60	14.05	49.23	38.19 ± 5.13	25.41	51.99	13.44	48.80
KNPS	55.44 ± 8.04	33.17	94.23	14.50	44.27	58.18 ± 7.93	36.60	95.40	13.63	38.76	51.34 ± 5.33	40.75	68.75	10.38	46.48
GNAS1	1.38 ± 0.35	0.60	3.13	25.36	35.51	1.45 ± 0.36	0.35	2.55	24.83	37.06	1.39 ± 0.27	0.70	2.15	19.42	34.34
GNAS2	1.48 ± 0.32	0.79	2.30	21.62	36.98	1.55 ± 0.35	0.65	2.45	22.58	38.46	1.64 ± 0.25	0.95	2.20	15.24	38.56
GNAS3	1.74 ± 0.34	0.90	2.80	19.54	37.94	1.78 ± 0.36	0.85	3.00	20.22	33.54	1.86 ± 0.20	1.10	2.40	10.75	31.73
GNAS	4.62 ± 0.94	2.51	7.40	20.35	38.65	4.82 ± 0.97	2.10	7.95	20.12	37.40	4.90 ± 0.67	2.95	6.75	13.67	43.62
GNBS1	0.28 ± 0.46	0.00	2.80	164.29	40.15	0.32 ± 0.46	0.00	2.80	143.75	36.65	0.34 ± 0.37	0.00	2.10	108.82	30.34
GNBS2	1.21 ± 0.84	0.00	4.00	69.42	32.79	1.28 ± 0.82	0.00	4.00	64.06	30.00	1.34 ± 0.67	0.10	3.05	50.00	33.71
GNBS3	2.50 ± 0.74	0.00	4.30	29.60	33.66	2.61 ± 0.69	0.20	4.30	26.44	33.17	2.55 ± 0.56	0.70	3.65	21.96	35.82
GNBS	3.98 ± 1.85	0.00	11.10	46.48	35.38	4.12 ± 1.77	0.20	11.10	42.96	37.08	4.21 ± 1.48	0.80	8.80	35.15	37.12
Trait	15YZ					16XX					BLUP				
	Mean ± SD	Min	Max	CV (%)	h^2	Mean ± SD	Min	Max	CV (%)	h^2	Mean ± SD	Min	Max	CV (%)	h^2
SNPS	20.17 ± 1.59	16.60	24.05	7.86	51.98	20.81 ± 1.47	14.80	24.20	7.05	55.43	20.88 ± 1.27	16.26	24.22	6.06	53.44
TKW	38.16 ± 4.95	25.48	52.83	12.96	47.56	44.75 ± 4.82	33.38	58.29	10.76	48.42	39.67 ± 4.51	27.76	51.25	11.37	49.34
KNPS	54.51 ± 6.5	41.70	79.60	11.92	43.70	59.47 ± 9.09	35.80	85.80	15.29	47.85	55.80 ± 6.04	39.17	84.64	10.82	48.51
GNAS1	1.59 ± 0.35	0.70	2.65	22.01	30.54	1.45 ± 0.39	0.60	2.20	26.90	35.61	1.45 ± 0.23	0.88	2.11	15.86	30.27
GNAS2	1.73 ± 0.30	0.90	2.60	17.34	34.32	1.53 ± 0.38	0.40	2.40	24.84	37.47	1.58 ± 0.21	0.90	2.18	13.29	33.11
GNAS3	1.99 ± 0.30	1.00	3.15	15.08	35.60	1.74 ± 0.38	0.80	3.00	21.84	37.68	1.82 ± 0.21	1.18	2.57	11.54	35.59
GNAS	5.27 ± 0.88	2.85	8.15	16.70	34.06	4.73 ± 0.98	1.80	7.40	20.72	37.93	4.85 ± 0.62	2.97	6.86	12.78	39.68
GNBS1	0.40 ± 0.47	0.00	2.55	117.50	32.99	0.68 ± 0.75	0.00	2.80	110.29	32.04	0.41 ± 0.34	0.06	1.93	82.93	35.31
GNBS2	1.34 ± 0.74	0.00	3.50	55.22	36.77	1.97 ± 1.02	0.00	4.20	51.78	32.92	1.43 ± 0.57	0.24	2.96	39.86	34.85
GNBS3	2.45 ± 0.65	0.60	4.10	26.53	39.73	2.98 ± 0.76	0.00	4.80	25.50	34.05	2.62 ± 0.48	0.72	3.64	18.32	37.60
GNBS	4.19 ± 1.74	0.75	10.15	41.53	41.63	5.67 ± 2.32	0.00	10.80	40.92	40.33	4.45 ± 1.29	1.04	8.38	28.99	36.99

SD standard deviation, *CV* coefficient of variation, h^2 board sense heritability

three basal spikelets (GNBS). Grain numbers in the three apical spikelets were designated as GNAS1, GNAS2, and GNAS3 from the apex downwards, and three basal spikelets were designated as GNBS1, GNBS2, and GNBS3 from the base upwards.

Genotyping and statistical analysis

Genomic DNA extraction was carried out according to CTAB method (Sharp et al. 1989). Descriptive statistical analysis and analysis of variance (ANOVA) of phenotypic data and G × E interaction were calculated by using SAS 9.4 (https://www.sas.com/en_us/software/sas9.html). The best linear unbiased prediction (BLUP) method was used to calculate the mean values of each trait (Bernardo 1996a, b; Bernardo et al. 1996). The broad sense heritability (h^2) was calculated according to the formula $h^2 = \sigma_g{}^2/(\sigma_g{}^2 + \sigma_e{}^2)$, where $\sigma_g{}^2$ is genetic variance and $\sigma_e{}^2$ is the residual variance.

SNP genotyping was performed on the BeadStation and iScan instruments and conducted at the Genome Center of the University of California at Davis according to the manufacturer's protocols (Illumina, USA) (Cavanagh et al. 2013). Data correction, input, and output were performed using GenomeStudio v2011.1 (Wang et al. 2014). Information on chromosome location of polymorphic SNPs was obtained from Cavanagh et al. (2013).

PowerMarker V3.25 was used to estimate genetic diversity of SNPs (Liu and Muse 2005). Population structure of the 212 cultivars was evaluated with 3792 SNP markers distributed on all 21 chromosomes using Structure 2.3.4 (Pritchard et al. 2000). The subpopulation number was estimated using the ΔK model (Evanno et al. 2005).

The average data from five environments were used for GWAS. The unified mixed model approach ($Q + K$ model) was applied to the data using TASSEL 5.0 to estimate marker-trait associations (MTAs) (Yu et al. 2005; Bradbury et al. 2007; Zhang et al. 2010). After exclusion of SNP loci with frequencies < 0.05, a uniform suggestive genome-wide significance threshold ($1/3271 = 3.06 \times 10^{-4}$, or $P < 3.06 \times 10^{-4}$, -Log$P > 3.51$) was given.

Markers with significant association with the traits were converted to 1 (favorable allele) and 0 (unfavorable allele) and were used for regression analysis. Trait values of different genotypes were predicted with these markers and compared with the actual values.

Results

Phenotypic assessment

All eleven traits (SNPS, TKW, KNPS, GNAS1, GNAS2, GNAS3, GNAS, GNBS1, GNBS2, GNBS3, and GNBS) were assessed in five environments (14JZ, 14YZ, 15JZ, 15YZ, and 16XX). The average coefficients of variation for these traits ranged from 6.06 to 164.29%, indicating that grain set in the materials was significantly affected by environments, especially GNBS1. The mean values of GNAS1, GNAS2, GNAS3, and GNAS across the five environments were 1.45, 1.58, 1.82, and 4.85, respectively. All three apical spikelets showed similar fertilities with the uppermost spikelets showing slightly less fertilities. In contrast, the three basal spikelets showed much greater differences in

Table 2 Analysis of variance of 11 traits in five environments of 212 wheat cultivars

Source of variation	SNPS		TKW		KNPS		GNAS1		GNAS2		GNAS3	
	df	SS	df	SS	df	SS	df	SS	df	SS	df	SS
Genotypes	211	5025.4***	211	64,902.9***	211	117,979.2***	211	230.7***	211	190.0***	211	196.7***
Environments	4	817.5***	4	17,722.1***	4	21,324.8***	4	16.6***	4	23.1***	4	27.8***
G × E interaction	844	1332.8***	844	13,516.2***	844	51,837.7***	844	143.5***	844	125.4***	844	123.4***
Source of variation	GNAS		GNBS1		GNBS2		GNBS3		GNBS			
	df	SS	df	SS	df	SS	df	SS	df	SS		
Genotypes	211	1608.9***	211	503.2***	211	1340.1***	211	961.3***	211	7065.9***		
Environments	4	147.8***	4	58.2***	4	230.0***	4	111.5***	4	1095.0***		
G × E interaction	844	779.7***	844	329.4***	844	742.9***	844	507.2***	844	3100.0***		

*** Indicate significance level at $P < 0.001$

Table 3 Pearson's correlation coefficients between phenotypic traits in the BLUPs

	TKW	KNPS	GNAS1	GNAS2	GNAS3	GNAS	GNBS1	GNBS2	GNBS3	GNBS
SNPS	− 0.067	0.356**	− 0.256**	− 0.270**	− 0.355**	− 0.309**	− 0.268**	− 0.231**	− 0.249**	− 0.264**
TKW		− 0.344**	− 0.201	− 0.220	− 0.230	− 0.228	− 0.115	− 0.125	− 0.134	− 0.135
KNPS			0.417**	0.398**	0.406**	0.506**	0.424**	0.455**	0.556**	0.590**
GNAS1				0.837**	0.845**	0.946**	0.344**	0.266**	0.320**	0.326**
GNAS2					0.871**	0.949**	0.347**	0.264**	0.266**	0.306**
GNAS3						0.952**	0.395**	0.283**	0.290**	0.336**
GNAS							0.381**	0.286**	0.309**	0.340**
GNBS1								0.829**	0.671**	0.876**
GNBS2									0.861**	0.975**
GNBS3										0.923**

** Significant at $P < 0.01$

floret fertilities with the average grain number for GNBS1, GNBS2, and GNBS3 being 0.41, 1.43, and 2.62, respectively (Table 1).

Table 1 shows that SNPS showed the highest h^2 (50.52–55.43%), followed by TKW (47.56–49.34%). Among the traits associated with grain setting, KNPS had the highest h^2 (38.76–48.51%). The heritability for grain number of apical and basal spikelets was 30.27–43.62% and 30.00–41.63%, respectively. Significant differences ($P < 0.0001$) were found among genotypes

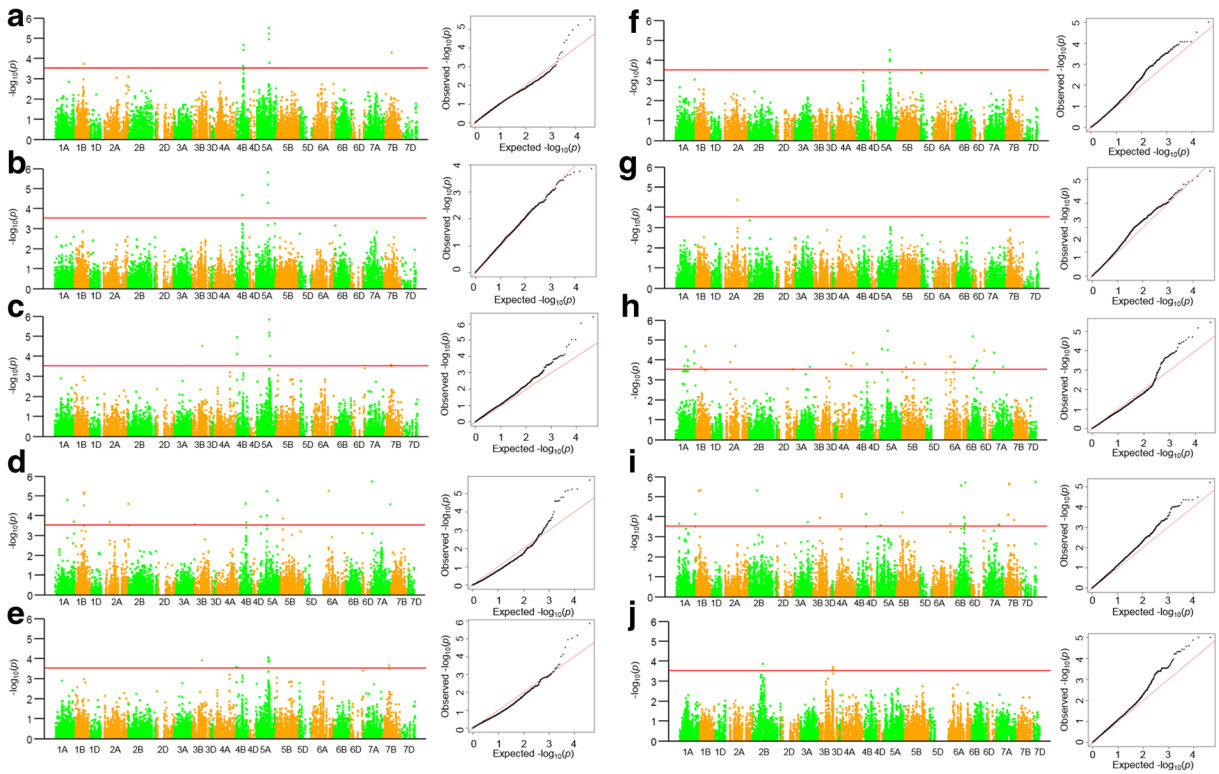

Fig. 1 Manhattan and Q-Q plots of eight phenotypic traits with 3778 genome-wide SNP markers shown as dot plots of compressed MLM at $P < 3.06 \times 10^{-4}$. Red horizontal line corresponds to the threshold value for significant association. Green and orange colors separate different chromosomes. **a** GNAS1. **b** GNAS2. **c** GNAS3. **d** GNAS. **e** GNBS1. **f** GNBS2. **g** GNBS3. **h** GNBS. **i** KNPS. **j** TKW

Table 4 Stable MTAs and phenotypic effects of favorable alleles revealed by GWAS consistently identified in at least two environments

Trait	SNP name	Chr.	cM	Favorable alleles	Freq. (%)	Environ.	Allele effect	*P* value	R^2
TKW	*wsnp_Ex_c3130_5790163*	3B	132.09	AA	75.00	16XX	0.57	2.03×10^{-4}	9.12
						BLUP	0.67	2.23×10^{-4}	10.42
KNPS	*wsnp_Ku_rep_c101175_88380491*	1A	36.81	CC	29.08	14JZ	3.02	4.52×10^{-6}	12.22
						14YZ	3.64	4.78×10^{-6}	14.88
	wsnp_Ex_c2025_3799847	1B	48.22	GG	32.19	15JZ	2.20	6.00×10^{-5}	8.48
						16XX	2.15	7.27×10^{-5}	7.90
	wsnp_Ex_c20786_29875033	2B	76.37	TT	39.22	14YZ	3.06	1.00×10^{-5}	8.85
						BLUP	2.10	1.18×10^{-5}	7.46
	wsnp_JD_c43971_30568640	3A	123.35	TT	37.75	14YZ	3.21	5.00×10^{-6}	14.47
						15YZ	3.00	6.15×10^{-6}	12.70
	wsnp_Ex_c6065_10623213	3B	66.39	AC	30.05	14JZ	2.95	5.30×10^{-6}	15.96
						16XX	3.05	7.52×10^{-6}	12.50
	wsnp_Ex_c28092_37240192	4A	140.47	GG	90.82	16XX	1.20	1.03×10^{-5}	8.17
						BLUP	0.24	1.52×10^{-5}	6.60
	wsnp_Ex_c5731_10066430	6B	84.83	CC	40.00	16XX	2.67	1.02×10^{-4}	7.20
						BLUP	1.85	1.30×10^{-4}	7.44
	wsnp_Ku_rep_c101817_88911480	6B	94.68	CC	61.38	16XX	1.31	1.10×10^{-5}	8.10
						BLUP	1.79	2.08×10^{-5}	7.36
	wsnp_Ra_c35321_43882919	7A	143.91	GG	44.30	15JZ	2.39	2.18×10^{-4}	6.54
						15YZ	1.80	2.40×10^{-4}	6.92
	wsnp_BF200891B_Ta_2_1	7B	40.62	TC	35.64	14JZ	2.52	4.10×10^{-5}	8.00
						15JZ	3.10	5.12×10^{-5}	7.20
	wsnp_Ex_c1790_3378771	7B	50.22	GG	35.42	15YZ	2.19	1.26×10^{-4}	6.50
						BLUP	1.35	1.85×10^{-4}	7.10
GNAS1	*wsnp_Ex_c32500_41144083*	4B	71.98	GG	14.15	15YZ	0.17	2.20×10^{-4}	8.56
						16XX	0.20	2.80×10^{-4}	7.10
	wsnp_Ex_c31799_40545376	5A	69.07	TT	7.11	14JZ	0.21	1.76×10^{-6}	18.18
						15JZ	0.26	1.76×10^{-6}	12.01
GNAS2	*wsnp_Ex_c31799_40545376*	5A	69.07	TT	7.11	14JZ	0.37	1.76×10^{-6}	13.20
						16XX	0.31	1.76×10^{-6}	10.16
GNAS	*wsnp_BG606986B_Ta_2_1*	1B	86.13	CC	50.00	14YZ	0.16	3.00×10^{-4}	6.30
						BLUP	0.14	3.40×10^{-4}	7.84
	wsnp_Ex_c31799_40545376	5A	69.07	TT	7.11	14YZ	1.36	1.00×10^{-4}	12.46
						BLUP	0.69	3.10×10^{-4}	9.91
GNBS1	*wsnp_Ra_c9755_16199734*	4B	14.80	TT	93.50	15JZ	0.11	1.03×10^{-4}	8.10
						15YZ	0.17	2.23×10^{-4}	7.38
	wsnp_BF293620A_Ta_2_3	5A	64.81	CC	12.43	14JZ	0.46	1.08×10^{-4}	13.50
						15JZ	0.33	1.82×10^{-4}	10.08
						16XX	0.39	2.33×10^{-4}	16.73
GNBS2	*wsnp_BF293620A_Ta_2_3*	5A	64.81	CC	12.43	14YZ	0.20	1.43×10^{-4}	11.20
						15YZ	0.23	1.62×10^{-4}	11.60
						BLUP	0.10	1.73×10^{-4}	12.10
GNBS	*wsnp_RFL_Contig2550_2175772*	1A	173.72	AA	88.61	15JZ	0.28	1.51×10^{-4}	6.37
						15YZ	0.10	3.77×10^{-5}	7.08

Table 4 (continued)

Trait	SNP name	Chr.	cM	Favorable alleles	Freq. (%)	Environ.	Allele effect	*P* value	R^2
	wsnp_BF293620A_Ta_2_3	5A	64.81	CC	12.43	16XX	0.41	2.03×10^{-4}	6.24
						BLUP	0.94	2.14×10^{-4}	6.31

(G), environments (E) for all 11 phenotype traits. G × E interactions were also significant (Table 2).

KNPS showed significant correlations ($P < 0.01$) with both GNAS and GNBS. Significant positive correlations were also found between GNAS and GNBS. However, SNPS showed a negative correlation with GNAS and GNBS. TKW showed insignificant negative correlations with all other traits but only significant with KNPS (Table 3).

Allelic diversity and genetic structure

Genotyping of the 212 wheat cultivars using the 9K SNP array identified 3778 polymorphic SNPs. Among them, 1793 were in the A genome chromosomes, 1778 were in the B genome, and 207 were in the D genome (Table S2). The values of gene diversity and polymorphism information content (PIC) ranged from 0.009 to 0.500 and from 0.009 to 0.375, with averages of 0.318 and 0.255, respectively. Major allele frequencies ranged up to 0.995 with an average of 0.765 (Table S2), indicating that the germplasm was highly diverse.

The number of subpopulation (K) was plotted against the ΔK calculated from the Structure, and the peak of the broken line graph was observed at $K = 2$ (Fig. S1a, b), indicating that the population was basically divided into two subpopulations.

GWAS of grain set-related traits and their phenotypic effects

Of the 3778 SNP markers, 3269 had frequencies above 0.05. Association analyses between the 11 traits and SNP markers showed that there were 112 significant associations ($P < 3.06 \times 10^{-4}$), with 4, 32, 33, and 43 for TKW, KNPS, apical, and basal grain set numbers, respectively (Fig. 1, Table S3). The associated loci were distributed on all chromosomes except 1D, 3D, 4D, and 5D (Table S3). Twenty-two SNP loci were significantly associated in at least two environments with phenotypic explanation rates (R^2) ranging from 6.24 to 18.18% (Table 4). Frequencies of favorable alleles at these associated loci ranged from 7.11 to 93.50%.

Most of GNAS-associated loci were distributed on six chromosomes, two on 5A (*wsnp_Ex_c2702_5013188* and *wsnp_Ex_c31799_40545376*), one on 4B (*wsnp_Ex_c32500_41144083*), and one on 2A (*wsnp_Ex_rep_c103167_88181968*). Two other loci with minor genetic effect were on 3B and 5B, respectively (Table S3). These six loci determined more than 30% phenotypic variation. The total GNAS predicted from these six markers showed very significant correlation with the actual numbers (Fig. 2a) with an increased number of favorable alleles increasing the total number of GNAS (Fig. 2b).

The SNP *wsnp_BF293620A_Ta_2_3* on 5A showed the largest effects on grain numbers in basal spikelets. Three other minor QTL were found on 1A, 2A, and 6B, respectively (Table S3). Similarly, the total GNBS predicted from these four markers showed a significant correlation with the actual GNBS (Fig. 3a) and an increased number of favorable alleles increased the total number of GNBS (Fig. 3b).

Stable SNPs for GNAS and GNBS

Two in 22 stable MTAs were significantly associated with two or more grain setting properties under various environmental conditions, including *wsnp_Ex_c31799_40545376-5A*$_{TT}$ (GNAS1, GNAS2, and GNAS) and *wsnp_BF293620A_Ta_2_3-5A*$_{CC}$ (GNBS1, GNBS2, and GNBS) (Table 4, Fig. 4a). The frequencies of favorable alleles of *wsnp_Ex_c31799_40545376-5A*$_{TT}$ (GNAS1, GNAS2, and GNAS) and *wsnp_BF293620A_Ta_2_3-5A*$_{CC}$ were 7.11% and 12.43%, respectively (Fig. 4b). The favorable allele at SNP locus *wsnp_Ex_c31799_40545376-5A* in 14JZ and 15JZ improved the grain set in the first apical spikelet by 0.21 and 0.26, increased the grain set of the second apical spikelet in 14JZ and 16XX by 0.37 and 0.31, and increased the grain set of the top three apical spikelets in 14YZ and BLUP by 1.36 and 0.69,

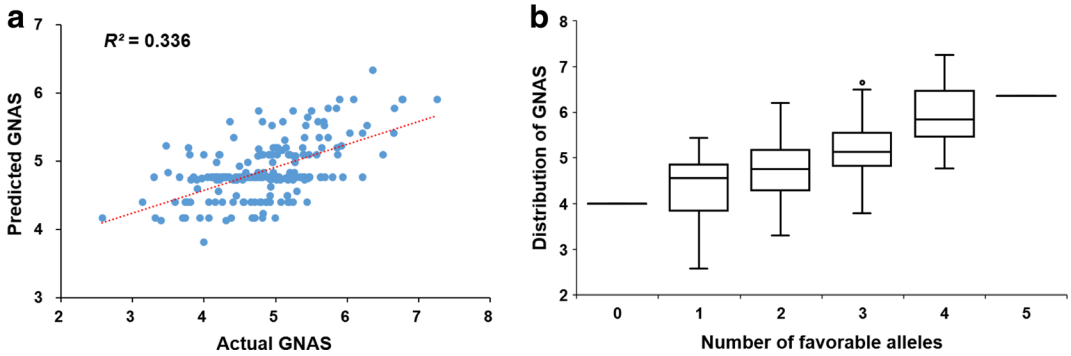

Fig. 2 Linear regression of predicted and actual numbers of GNAS (**a**); boxplot among the 212 wheat cultivars in BLUP between number of favored alleles and GNAS (**b**). The predicted GNAS are calculated from the following equation: $y = 0.82x1 + 0.58x2 + 0.60x3 + 0.36x4 + 0.43x5 + 3.80$, where y is the predicted GNAS; x1 is *wsnp_Ex_c2702_5013188*; x2 is *wsnp_CAP7_c1405_706142*; x3 is *wsnp_Ku_c35386_44598937*; x4 is *wsnp_Ex_rep_c68599_67447880*; and x5 is *wsnp_Ex_c3130_5789791*

respectively (Fig. 4c). Among the favorable alleles for basal grain setting, *wsnp_BF293620A_Ta_2_3-5A$_{CC}$* increased the grain sets of the first [0.46 (14JZ), 0.33 (15JZ), 0.39 (16XX)], the second [0.20 (14YZ), 0.23 (15YZ), 0.10 (BLUP)], and the bottom three [0.41 (16XX), 0.94 (BLUP)] basal spikelets (Table 4, Fig. 4d).

Discussion

KNPS is a fundamental yield component comprised of apical, basal, and middle spikelets. Assuming all other yield-determining factors are fixed, an increase in grain set in apical and basal spikelets could modestly but significantly improve the yield of wheat (Arisnabarreta and Miralles 2006; Acreche et al. 2008). However, not enough effort has been made to improve the fertility of apical and basal spikelets in breeding program. Both GNAS and GNBS were positively correlated with KNPS ($P < 0.01$) with only weak but insignificant negative correlations with TKW (Table 3). Our study also showed a large variation in both GNAS (2.1–7.1) and GNBS (0.5–9.2), indicating a great potential for improving the fertility of both basal and apical spikelets, thus increasing the total number of grains per spike.

Many MTAs for GNAS and GNBS were identified in this study. Most of them are in similar positions to those for grain yield and yield components (Table 5). The *wsnp_Ku_rep_c68318_67259259* for GNAS1 on chromosome 4B was associated with grain yield (Ain et al. 2015). The SNPs *wsnp_RFL_Contig4134_4692458* and *wsnp_Ex_c2288_4293430* associated with GNBS on chromosome 2D and 4A were mapped to QTL interval *Kukri_c14902_1112–RAC875_c77816_365* and *Kukri_rep_c106490_583–RAC875_c29282_566* that affected KNPS (Gao et al. 2015). A QTL for spike

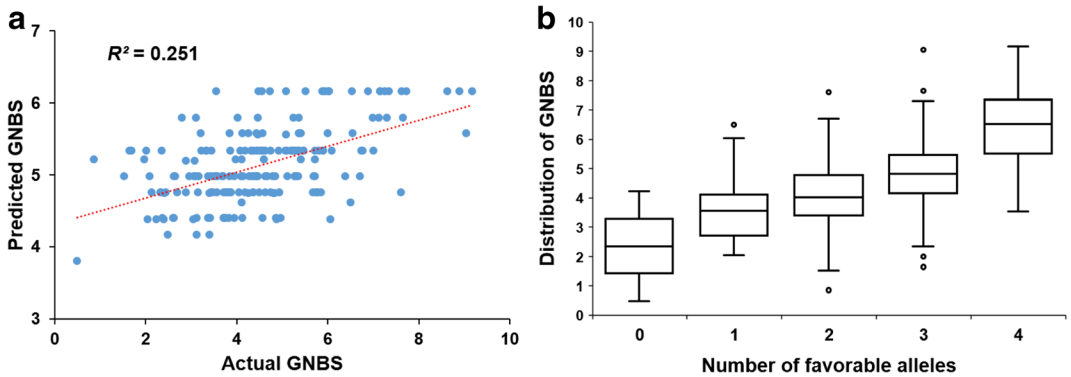

Fig. 3 Linear regression of predicted and actual numbers of GNBS (**a**); boxplot among the 212 wheat cultivars in BLUP between number of favored alleles and GNBS (**b**). The predicted GNBS are calculated from the following equation: $y = 1.25x1 + 0.76x2 + 0.93x3 + 0.67x4 + 2.45$, where y is the predicted GNBS; x1 is *wsnp_Ex_c54193_57155632*; x2 is *wsnp_Ra_c4850_8698731*; x3 is *wsnp_Ex_c15595_23910900*; and x4 is *wsnp_Ex_c8588_14419007*

number/m^2 (SN) in the marker interval *BS00032003_51-BS00070871_51* (Gao et al. 2015) was in a similar position to *wsnp_Ex_c607_1204733* which was found to be associated with GNBS in this study. The SNP *wsnp_Ex_c11446_18468102* associated with GNBS on chromosome 6A was located in a pleiotropic region, affecting TKW and SN (Gao et al. 2015). Another SNP marker *wsnp_Ex_c32500_41144083* associated with GNAS1, GNAS2, and GNAS on chromosome 4B was close to QTL interval (*IWB67166–IWB25207*) that affected days to maturity (Milner et al. 2016). However, its physical location on chromosome 4B is 574.9 Mb, which is far from the *Rht-B1* gene (30.8 Mb) (Table S4).

Gene pyramiding has been proved to be an effective approach in improving not only a plant's tolerance to biotic stresses (Zheng et al. 2017) and abiotic stresses (Zhou 2011) but also other agronomic traits and yield components (Mirabella et al. 2015). Among all MTAs for different traits, the SNPs *wsnp_Ex_c31799_40545376*

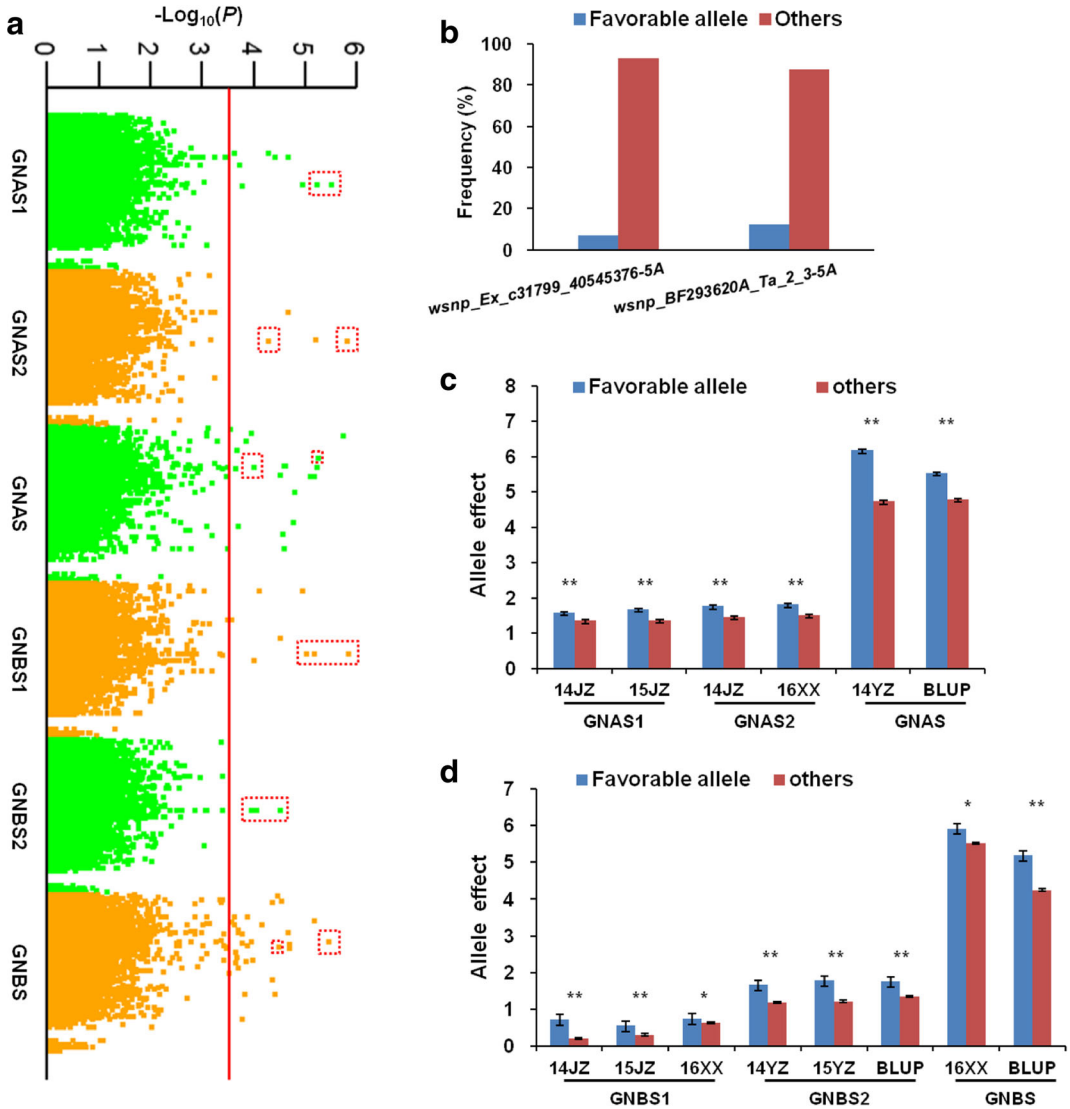

Fig. 4 Effect of favorable allele *wsnp_Ex_c31799_40545376-5A* and *wsnp_BF293620A_Ta_2_3-5A*. **a** Associated loci identified in the germplasm set by a mixed linear model ($P < 3.06 \times 10^{-4}$); green and orange colors separate different traits. **b** Allelic frequencies of *wsnp_Ex_c31799_40545376-5A* and *wsnp_BF293620A_Ta_2_3-5A* in the germplasm set; blue columns represent favorable alleles TT and CC, respectively. **c** Genetic effects of *wsnp_Ex_c31799_40545376-5A$_{TT}$* in selected environments. **d** Genetic effects of *wsnp_BF293620A_Ta_2_3-5Acc* in selected environments

Table 5 Significant SNP loci identified in current and previous studies

Trait	Chr.	Identified loci in current study	Near locus previously reported in the same chromosome
KNPS	1B	*wsnp_Ex_c2025_3799847*	*Excalibur_c8052_541* (Ain et al. 2015)
	2B	*wsnp_Ex_c20786_29875033*	*Excalibur_c19260_105—IACX8096* (Gao et al. 2015)
	7A	*wsnp_Ra_c35321_43882919*	*TA006231-0789* (Sun et al. 2017)
	7B	*wsnp_BF200891B_Ta_2_1*	*BS00011652_51—BS00081132_51* (Gao et al. 2015)
	7B	*wsnp_Ex_c1790_3378771*	*IWA2568–IWA6901* (Milner et al. 2016, durum wheat)
GNAS1	4B	*wsnp_Ku_rep_c68318_67259259*	*RAC875_c23144_1560* (Ain et al. 2015)
	4B	*wsnp_Ex_c32500_41144083*	*IWB67166-IWB25207* (Milner et al. 2016, durum wheat)
GNAS2	4B	*wsnp_Ex_c32500_41144083*	*IWB67166-IWB25207* (Milner et al. 2016, durum wheat)
GNAS	4B	*wsnp_Ex_c32500_41144083*	*IWB67166-IWB25207* (Milner et al. 2016, durum wheat)
GNBS	2D	*wsnp_RFL_Contig4134_4692458*	*Kukri_c14902_1112-RAC875_c77816_365* (Gao et al. 2015)
	4A	*wsnp_Ex_c2288_4293430*	*Kukri_rep_c106490_583-RAC875_c29282_566* (Gao et al. 2015)
	5B	*wsnp_Ex_c607_1204733*	*BS00032003_51-BS00070871_51* (Gao et al. 2015)
	6A	*wsnp_Ex_c11446_18468102*	*Ku_c32392_967-wsnp_RFL_Contig2523_2130662* (Gao et al. 2015)

and *wsnp_BF293620A_Ta_2_3* on 5A showed consistent significant association with all grain setting properties under various environmental conditions (Table 4, Fig. 2). The corresponding overlapping genes related to these two loci have not been reported for either spikelet fertility or yield-related genes according to the published sequence of the hexaploid wheat genome (Table S4, Fig. S2). Interestingly, *Vrn-B* gene was found in the 4.5-Mb region, and its physical distance to *wsnp_Ex_c31799_40545376* and *wsnp_BF293620A_Ta_2_3* was 2 Mb and 1.1 Mb, respectively. Therefore, the new markers can be potentially used in breeding programs to improve the fertility of both basal and apical spikelets. Further studies should be conducted using a segregating population to identify the gene and verify their roles in spikelet fertility in wheat.

In our study, apart from the MTAs on 5A, several other MTAs were identified for GNAS and GNBS. The combination of the favorable alleles from different MTAs significantly improved the fertility of both GNAS and GNBS (Figs. 2 and 3). From the 212 wheat accessions used in this study, less than 10% of accessions have favorable allele on 5A and just around 10% of accessions have favorable allele for the other significant MTA on 4B, suggesting that less effort has been made in improving the fertility of GNAS and GNBS. No accessions were found to have favorable alleles from both significant MTAs, i.e., 5A and 4B MTAs, which opens the door for breeders to improve the fertility of GNAS and GNBS by pyramiding those two loci.

In conclusion, two new MTAs were identified for the fertility of basal and apical spikelets, respectively. Both of loci were located on chromosome 5A and not found to be associated with any grain setting properties in previous studies. The combination of these loci with other MTAs for spikelet fertility could improve the grain setting in both basal and apical spikelets.

Funding information This work was supported by grants from the National Key R&D Program of China (2017YFD0101000), the National Key R&D Program of Shanxi Province (201703D211007), the Technology Innovation Program of Higher Education of Shanxi Province (2017142), and the Science & Technology Innovation Foundation of Shanxi Agricultural University (2016YJ05).

References

Acreche MM, Briceño-Félix G, Sánchez JAM, Slafer GA (2008) Physiological bases of genetic gains in Mediterranean bread wheat yield in Spain. Eur J Agron 28:162–170

Ain Q, Rasheed A, Anwar A, Mahmood T, Imtiaz M, Mahmood T, Xia X, He Z (2015) Genome-wide association for grain yield under rainfed conditions in historical wheat cultivars from Pakistan. Front Plant Sci 6:743

Akter MB, Piao R, Kim B, Lee Y, Koh E, Koh HJ (2014) Fine mapping and candidate gene analysis of a new mutant gene for panicle apical abortion in rice. Euphytica 197:387–398

Arisnabarreta S, Miralles DJ (2006) Floret development and grain setting in near isogenic two- and six-rowed barley lines (*Hordeum vulgare* L.). Field Crop Res 96:466–476

Bancal P (2009) Early development and enlargement of wheat floret primordia suggest a role of partitioning within spike to grain set. Field Crop Res 110:44–53

Bernardo R (1996a) Test cross additive and dominance effects in best linear unbiased prediction of maize single-cross performance. Theor Appl Genet 93:1098–1102

Bernardo R (1996b) Best linear unbiased prediction of maize single-cross performance. Crop Sci 36:50–56

Bernardo R, Murigneux A, Karaman Z (1996) Marker-based estimates of identity by descent and alikeness in state among maize inbreds. Theor Appl Genet 93:262–267

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

Cavanagh CR, Chao S, Wang S, Huang BE, Stephen S, Kiani S, Forrest K, Saintenac C, Brown-Guedira GL, Akhunova A, See D, Bai G, Pumphrey M, Tomar L, Wong D, Kong S, Reynolds M, da Silva ML, Bockelman H, Talbert L, Anderson JA, Dreisigacker S, Baenziger S, Carter A, Korzun V, Morrell PL, Dubcovsky J, Morell MK, Sorrells ME, Hayden MJ, Akhunov E (2013) Genome-wide comparative diversity uncovers multiple targets of selection for improvement in hexaploid wheat landraces and cultivars. Proc Natl Acad Sci U S A 110:8057–8062

Cheng ZJ, Mao BG, Gao SW, Zhang L, Wang JL, Lei CL, Zhang X, Wu FQ, Guo XP, Wan J (2011) Fine mapping of *qPAA8*, a gene controlling panicle apical development in rice. J Integr Plant Biol 53:710–718

Dencic S, Kastori R, Kobiljski B, Duggan B (2000) Evaluation of grain yield and its components in wheat cultivars and landraces under near optimal and drought conditions. Euphytica 113:43–52

Dreccer MF, Wockner KB, Palta JA, McIntyre CL, Borgognone MG, Bourgault M, Reynolds M, Miralles DJ (2014) More fertile florets and grains per spike can be achieved at higher temperature in wheat lines with high spike biomass and sugar content at booting. Funct Plant Biol 41:482

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

Ferrante A, Savin R, Slafer GA (2010) Floret development of durum wheat in response to nitrogen availability. J Exp Bot 61:4351–4359

Ferrante A, Savin R, Slafer GA (2012) Differences in yield physiology between modern, well adapted durum wheat, cultivars grown under contrasting conditions. Field Crop Res 136:52–64

Ferrante A, Savin R, Slafer GA (2013a) Floret development and grain setting differences between modern durum wheats under contrasting nitrogen availability. J Exp Bot 64:169–184

Ferrante A, Savin R, Slafer GA (2013b) Is floret primordia death triggered by floret development in durum wheat? J Exp Bot 64:2859–2869

Fischer RA (2008) The importance of grain or kernel number in wheat: a reply to Sinclair and Jamieson. Field Crop Res 105: 15–21

Gallavotti A, Malcomber S, Gaines C, Stanfield S, Whipple C, Kellogg E, Schmidt RJ (2011) BARREN STALK FASTIGIATE1 is an AT-hook protein required for the formation of maize ears. Plant Cell 23:1756–1771

Gao FM, Wen WE, Liu JD, Rasheed A, Yin GH, Xia XC, Wu XX, He ZH (2015) Genome-wide linkage mapping of QTL for yield components, plant height and yield-related physiological traits in the Chinese wheat cross Zhou 8425B/Chinese spring. Front Plant Sci 6:1099

Gao FM, Ma DY, Yin GH, Rasheed A, Dong Y, Xiao YG, Xia XC, Wu XX, He ZH (2017) Genetic progress in grain yield and physiological traits in Chinese wheat cultivars of Southern Yellow and Huai valley since 1950. Crop Sci 57:760–773

González FG, Miralles DJ, Slafer GA (2011) Wheat floret survival as related to pre-anthesis spike growth. J Exp Bot 62:4889–4901

González-Navarro OE, Griffiths S, Molero G, Reynolds MP, Slafer GA (2015) Dynamics of floret development determining differences in spike fertility in an elite population of wheat. Field Crop Res 172:21–31

Guo Z, Schnurbusch T (2015) Variation of floret fertility in hexaploid wheat revealed by tiller removal. J Exp Bot 66:5945–5958

Guo J, Zhang Y, Shi WP, Zhang BQ, Zhang JJ, Xu YH, Cheng XM, Cheng K, Zhang XY, Hao CY, Cheng SH (2015) Front Plant Sci 6:1029. Association analysis of grain-setting rate at the apical and basal spikelets in bread wheat (*Triticum aestivum* L

Guo Z, Chen D, Alqudah AM, Röder MS, Ganal MW, Schnurbusch T (2017) Genome-wide association analyses of 54 traits identified multiple loci for the determination of floret fertility in wheat. New Phytol 214:257–270

Hawkesford MJ, Araus JL, Park R, Calderini D, Miralles D, Shen T, Zhang J, Parry MAJ (2013) Prospects of doubling global wheat yields. Food Energy Secur 2:34–48

Langer RHM, Hanif M (1973) A study of floret development in wheat (*Triticum aestivum* L.). Ann Bot 37:743–751

Li SB, Qian Q, Fu ZM, Zeng DL, Meng XB, Kyozuka J, Maekawa M, Zhu XD, Zhang J, Li JY, Wang YH (2009) *Short panicle1* encodes a putative PTR family transporter and determines rice panicle size. Plant J 58:592–605

Liu KJ, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21: 2128–2129

Meng ZD, Zhang FJ, Ding ZH, Sun Q, Wang LM, Guo QF, Wang HG (2007) Inheritance of ear tip-barrenness trait in maize. Agric Sci China 6:628–633

Milner SG, Maccaferri M, Huang BE, Mantovani P, Massi A, Frascaroli E, Tuberosa R, Salvi S (2016) A multiparental cross population for mapping QTL for agronomic traits in durum wheat (*Triticum turgidum* ssp. *durum*). Plant Biotechnol J 14:735–748

Mirabella NE, Abbate PE, Ramirez IA, Pontaroli AC (2015) Genetic variation for wheat spike fertility in cultivars and early breeding materials. J Agric Sci 154:13–22

Mishra SP, Mohapatra PK (1987) Soluble carbohydrates and floret fertility in wheat in relation to population density stress. Ann Bot 60:269–277

Piao S, Ciais P, Huang Y, Shen Z, Peng S, Li J, Zhou L, Liu H, Ma Y, Ding Y, Friedlingstein P, Liu C, Tan K, Yu Y, Zhang T,

Fang J (2010) The impacts of climate change on water resources and agriculture in China. Nature 467:43–51

Pritchard JK, Stephens M, Rosenberg NA, Donnelly P (2000) Association mapping in structured populations. Am J Hum Genet 67:170–181

Rerkasem B, Jamjod S (1997) Boron deficiency induced male sterility in wheat (*Triticum aestivum* L.) and implications for plant breeding. Euphytica 96:257–262

Reynolds M, Foulkes MJ, Slafer GA, Berry P, Parry MA, Snape JW, Angus WJ (2009) Raising yield potential in wheat. J Exp Bot 60:1899–1918

Saifuzzaman M, Fattah QA, Islam MS (2008) Spikelet sterility of wheat in farmer's field in Northwest Bangladesh. Bangladesh J Bot 37:155–160

Satoh-Nagasawa N, Nagasawa N, Malcomber S, Sakai H, Jackson D (2006) A trehalose metabolic enzyme controls inflorescence architecture in maize. Nature 441:227–230

Sharp PJ, Chao S, Desai S, Gale MD (1989) The isolation, characterization and application in the Triticeae of a set of wheat RFLP probes identifying each homoeologous chromosome arm. Theor Appl Genet 78:342–348

Shitsukawa N, Kinjo H, Takumi S, Murai K (2009) Heterochronic development of the floret meristem determines grain number per spikelet in diploid, tetraploid and hexaploid wheats. Ann Bot 104:243–251

Sun C, Zhang F, Yan X, Zhang X, Dong Z, Cui D, Chen F (2017) Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai Valley of China. Plant Biotechnol J 15:953–969

Tan CJ, Sun YJ, Xu HS, Yu SB (2011) Identification of quantitative trait locus and epistatic interaction for degenerated spikelets on the top of panicle in rice. Plant Breed 130:177–184

Tripathi A, Tripathi DK, Chauhan DK, Kumar N, Singh GS (2016) Paradigms of climate change impacts on some major food sources of the world: a review on current knowledge and future prospects. Agric Ecosyst Environ 216:356–373

Wang SC, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, International Wheat Genome Sequencing Consortium, Lillemo M, Mather D, Appels R, Dolferus R, Brown-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90000 single nucleotide polymorphism array. Plant Biotechnol J 12:787–796

Yamagishi J, Miyamoto N, Hirotsu S, Laza RC, Nemoto K (2004) QTLs for branching, floret formation, and pre-flowering floret abortion of rice panicles in a temperate japonica × tropical japonica cross. Theor Appl Genet 109:1555–1561

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang ZW, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360

Zheng C, Zhu Y, Zhu HJ, Kang G, Guo T, Wang C (2014) Floret development and grain setting characteristics in winter wheat in response to pre-anthesis foliar applications of 6-benzylaminopurine and boron. Field Crop Res 169:70–76

Zheng C, Zhu Y, Wang C, Guo T (2016) Wheat grain yield increase in response to pre-anthesis foliar application of 6-benzylaminopurine is dependent on floret development. PLoS One 11:e0156627

Zheng Z, Gao S, Zhou MX, Yan GJ, Liu CJ (2017) Enhancing Fusarium crown rot resistance by pyramiding large-effect QTL in common wheat (*Triticum aestivum* L.). Mol Breed 37:107

Zhou MX (2011) Accurate phenotyping reveals better QTL for waterlogging tolerance in barley. Plant Breed 130:203–208