

A genome-wide association study of wheat yield and quality-related traits in southwest China

Jing Liu · Bo Feng · Zhibin Xu · Xiaoli Fan ·
Fei Jiang · Xiufeng Jin · Jun Cao · Fang Wang ·
Qin Liu · Li Yang · Tao Wang

Received: 7 August 2017 / Accepted: 20 November 2017 / Published online: 1 December 2017
© Springer Science+Business Media B.V., part of Springer Nature 2017

Abstract Wheat (*Triticum aestivum* L.) is one of the most productive and important crops and its yield potential and quality characteristics are tightly linked with the global food security. In this study, genome-wide association study (GWAS) was conducted for yield and quality-related traits. Based on the high-density wheat 90K Illumina iSelect SNP Array, 192 bread wheat lines from southwest China, including 25 synthetic hexaploid wheat lines, 80 landraces, and 87 cultivars were analyzed. Association analysis results indicated that there were 57, 27, 30, and 34 SNPs associated with plant height (PH), grain protein content (GPC), thousand kernel weight (TKW), and SDS sedimentation volume (SSV) have been detected, respectively. Then, integrating RNA-Seq with bioinformatics analysis, 246 candidate genes (102 for GPC, 52 for TKW, and 92 for SSV) were found. Further analysis indicated that one up-regulated and two down-regulated expression genes

affect GPC. Additionally, two haplotypes significantly affecting PH were detected in a 2.2-Mb genome region encompassing a gene which encoded an ubiquitin-specific protease, TaUBP24. The functional markers of *TaUBP24* have been developed, which could be used for marker-assisted selection to improve wheat quality and yield.

Keywords Wheat · GWAS · Plant height · Grain protein content · Ubiquitin-specific protease

Abbreviations

ABA	Abscisic acid
CMLM	Compressed mixed linear model
CTAB	Cetyl trimethyl ammonium bromide
GA	Gibberellic acid
GPC	Grain protein content
GWAS	Genome-wide association study
HWE	Hardy-Weinberg equilibrium
KEGG	Kyoto Encyclopedia of Genes and Genomes
MAF	Minor allele frequency
PCA	Principal component analysis
PH	Plant height
QTL	Quantitative trait locus
RCBD	Randomized complete block design
SSV	SDS sedimentation volume
SNP	Single nucleotide polymorphism
TKW	Thousand kernel weight
UBP24	Ubiquitin-specific protease 24

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11032-017-0759-9>) contains supplementary material, which is available to authorized users.

J. Liu · B. Feng · Z. Xu · X. Fan · F. Jiang · X. Jin ·
J. Cao · F. Wang · Q. Liu · L. Yang · T. Wang (✉)
Chengdu Institute of Biology, Chinese Academy of Sciences,
Chengdu 610041, China
e-mail: wangtao@cib.ac.cn

J. Liu
University of Chinese Academy of Sciences, Beijing 100049,
China

Introduction

Wheat (*Triticum aestivum* L.) is a crucial cereal crop for the demand of global food security (Curtis and Halford 2014). The production of wheat already reached 751.1 million tons in 2016 (USDA 2017). However, wheat is now cultivated and consumed more widely and a lot of products such as breads, biscuits, noodles, cakes, and beer have been developed (Curtis and Halford 2014). The further improvement of the yield potential and quality traits of wheat so as to meet current and impending challenges is necessary.

It has been reported that a lot of important agronomic traits, such as plant height (PH), grain protein content (GPC), etc., are controlled by multiple genes that are often affected by various environmental conditions (Groos et al. 2003; Prasad et al. 2003). Generally, quantitative trait locus (QTL) mapping is used to detect these corresponding loci. For example, the loci for GPC have been performed in different populations (Distelfeld et al. 2008; Echeverry-Solarte et al. 2015; Li et al. 2012a; Nigro et al. 2014; Prasad et al. 2003; Terasawa et al. 2016). However, the traditional QTL mapping approach could only locate genomic regions with low resolution and be limited to the bi-parental population (Sukumaran et al. 2015).

Based on the test of correlations between genotype and phenotype in a set of individuals consist of germplasm collections or natural populations (Marcotuli et al. 2015), genome-wide association study (GWAS) is a powerful tool widely used in breeding programs for it could be more efficient and less laborious in analyzing complex traits under varied environmental conditions. For example, in rice, a high-density SNP (single nucleotide polymorphism) array was used for identifying grain length and width loci (Chen et al. 2014). In soybean, GWAS was carried out for oil content using the re-sequenced lines (Zhou et al. 2015). In maize, many beneficial alleles were identified for kernel starch, protein, and oil by using GWAS method (Cook et al. 2012). In the same way, there were a lot of successful cases of GWAS applied in wheat analysis, such as rust resistance (Gao et al. 2016; Maccaferri et al. 2015; Naruoka et al. 2015; Yu et al. 2011; Zegeye et al. 2014), Fusarium head blight resistance (Kollers et al. 2013), arabinoxylan content (Marcotuli et al. 2015), grain morphology (Bresseghele and Sorrells 2006;

Rasheed et al. 2014), and grain yield and related traits (Ain et al. 2015; Lopes et al. 2013; Sukumaran et al. 2015; Zanke et al. 2014a, b, 2015).

In this study, the main wheat cultivars and excellent germplasm collections were selected to conduct GWAS by using the high-density wheat 90 K Illumina iSelect SNP Array (Wang et al. 2014). The SNPs associated with yield and quality-related traits such as PH, GPC, thousand kernel weight (TKW), and SDS sedimentation volume (SSV) were identified which could ultimately facilitate breeding by whole genome selection.

Materials and methods

Genetic resources

A total of 192 bread wheat lines were used for this study including 25 synthetic hexaploid wheat lines, 80 landraces, and 87 cultivars (Table S1). Wheat plants were grown in two locations of Sichuan Province, in two growing seasons, 2014–2015 and 2015–2016. The locations were Shuangliu (N content 1.604 g kg⁻¹) and Shifang (N content 1.680 g kg⁻¹). High and low nitrogen treatment was also designed in 2015–2016 growing season in Shifang. The main cultural practices consisted in nitrogen application (60 kg N ha⁻¹) in the high nitrogen treatment place and without nitrogen application in the low nitrogen treatment place after seeding. The lines were sown in single row of 1.2 m long and 0.2 m apart following a randomized complete block design (RCBD) with two replications per location, and harvested by hand. Three plants of each line per replication were selected at random for phenotype analysis.

Genotyping and SNP calling

Total DNA was extracted from the fresh leaves of seedling wheat using the cetyl trimethyl ammonium bromide (CTAB) method (Murray and Thompson 1980), and genotyped with the wheat 90 K Illumina iSelect SNP Array at the Compass Biotechnology Co., Ltd. All SNPs were filtered according to the criteria: minor allele frequency (MAF) greater than 0.05, sample call rate more than 0.8 and SNP call rate more than 0.9, and Hardy-Weinberg equilibrium (HWE) value less than 0.000001.

Population genetics and linkage disequilibrium analysis

A phylogenetic tree was constructed using pruned SNPs by the neighbor-joining method and displayed in MEGA 6.06 (http://www.megasoftware.net/archived_version_active_download) (Tamura et al. 2013). Population structure (Q) was evaluated using the filtered SNPs by ADMIXTURE (<http://www.genetics.ucla.edu/software/admixture/download.html>) (Alexander et al. 2009). A principal component analysis (PCA) of filtered SNPs was performed with the GAPIT (<http://zzlab.net/GAPIT/>) (Lipka et al. 2012) and the first two PCA values were plotted in two dimensions. Linkage disequilibrium was estimated using PLINK (<http://www.cog-genomics.org/plink2>) (Chang et al. 2015) by estimating the squared allele frequency correlations R^2 value from pruned SNPs.

Phenotyping

PH was measured after-harvest, as the average height (cm) of each point from the ground part of the root to the tip of the spike, excluding the awns. TKW was determined from the weight of more than 200 random kernels with two technical repeats by SC-E software (Hangzhou Wanshen Detection Technology Co., Ltd., Hangzhou, China). Three grams of each variety was subsequently ground to whole meal using a 1-mm-sieve Cyclotec mill (Foss Tecator AB, höganäs, Sweden). GPC was determined twice for each line by a Kjeltac 2200 system (FOSS). The SSV was measured with two duplications according to AACC 56-61A criterion (<https://doi.org/10.1094/AACCIntMethod-56-61.02>) (AACC International 1999).

Genome-wide association study

A compressed mixed linear model (CMLM) program which took the results of population stratification and kinship as covariate was used for the association analysis to minimize false positives (Lipka et al. 2012). Also, the first three PCA values, which were derived from pruned SNPs, were used as covariate in the model to adjust population stratification. For GWAS results, a threshold P -value of 0.001 ($-\log_{10}P = 3$) was used to declare significant SNPs.

RNA sequencing and analysis

Twelve lines were selected for RNA-Seq analysis. Immature grain tissue was collected from plants at 14 and 21 day after flowering. RNA sequencing and data analysis were done by the Guangzhou Gene denovo Biotechnology Co., Ltd.

Multiple alignment and phylogenetic analysis

The gene structure was analyzed using the online software GSDS 2.0 (Hu et al. 2015). The amino acid sequences were aligned using the ClustalW (<http://www.clustal.org/download/current>) (Larkin et al. 2007). The homologous amino acid sequences from *Triticum aestivum*, *Arabidopsis thaliana*, *Hordeum vulgare* subsp. *vulgare*, *Brachypodium distachyon*, *Oryza sativa* spp. *japonica*, *Setaria italica*, *Zea mays*, *Sorghum bicolor*, *Glycine max* and *Homo sapiens* were used for phylogenetic tree construction. The neighbor-joining program with the parameters selection of P distance model, pairwise deletion, and 1000 bootstrap was conducted and displayed by using MEGA 6.06.

Results

Genotyping and population genetics analysis

A total of 192 bread wheat samples including synthetic hexaploid wheat lines, landraces, and cultivars were used in this study. SNPs calling of the wheat lines were generated by wheat 90 K Illumina iSelect SNP Array. The call rate was 0.977813. There were 13,154 polymorphic SNPs selected after quality control.

The phylogenetic relationships among the 192 wheat lines were analyzed through filtered SNPs. They were separated into three parts on the basis of synthetic hexaploid wheat lines (red line), landraces (green line), and cultivars (blue line), although there were a few exceptions (Fig. 1a). For instance, three landraces were closed to cultivars, one cultivar and three synthetic hexaploid wheat lines fell together with the landraces, and one landrace and one synthetic hexaploid wheat line got together with the cultivars as revealed by the phylogenetic tree. Although there was no obvious population structure among the

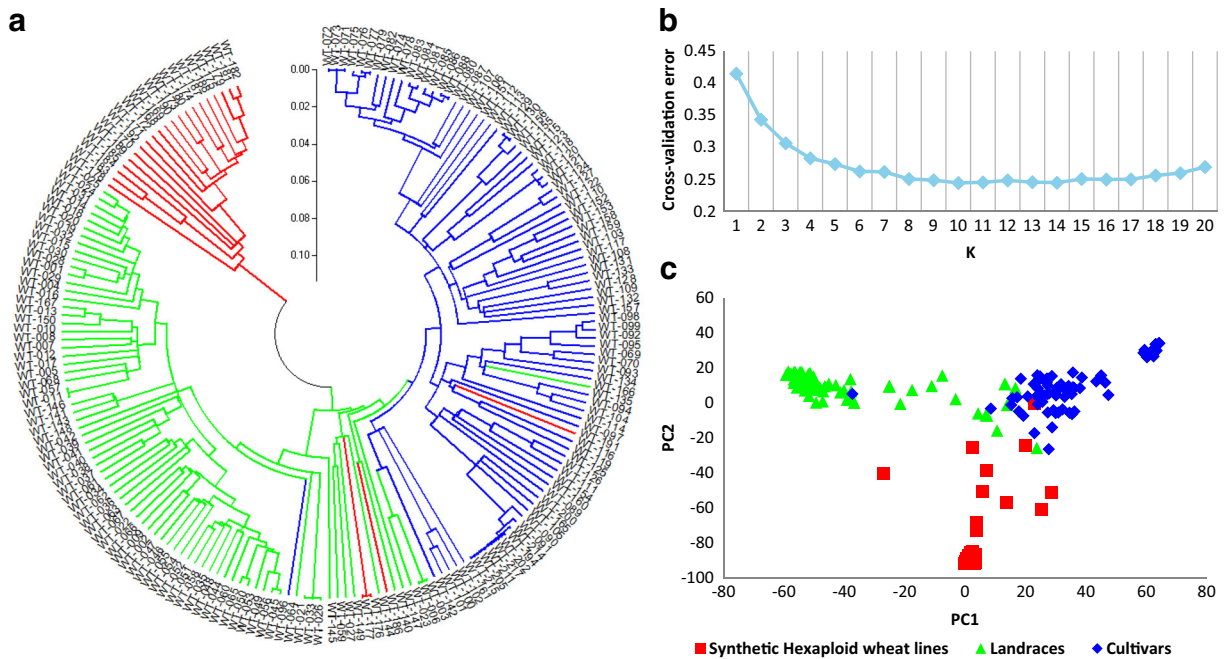


Fig. 1 Population genetic analysis of 192 wheat lines. **a** A neighbor-joining tree of the 192 wheat lines. **b** Population structure analysis of 192 wheat lines. The numbers of clusters (K) were

wheat lines (Fig. 1b), the selection, domestication, and improvement breeding process may have caused a hidden population structure (Neumann et al. 2010). So, the samples were divided into three groups according to synthetic hexaploid wheat lines, landraces, and cultivars, and this classification method was supported by the principal component analysis (Fig. 1c).

Phenotyping

The PH, TKW, and SSV in four environments and GPC in five environments were assayed and showed that there were broad variations for each character (Fig. S1). The mean value of PH was increased from the synthetic hexaploid wheat lines to the landraces (118.1 to 132.7 cm), but strong decrease was observed from the landraces to the cultivars (81.6 cm) with a reduction of more than 100 cm from the shortest to the tallest line (Fig. S2a, Table S1). In contrast with PH, the mean value of TKW firstly decreased from 45.0 g in the synthetic hexaploid wheat lines to 31.6 g in the landraces and then increased nearly the same to the previous level in the cultivars (44.6 g) (Fig. S2b, Table S1). However, the mean value of GPC gradually reduced from synthetic

set from 1 to 20. **c** PCA of the 192 wheat lines. Individuals from the same group were represented by the same color. Synthetic hexaploid wheat lines (red), landraces (green), and cultivars (blue)

hexaploid wheat lines (14.4%) to cultivars (11.7%) over time (Fig. S2c, Table S1). In comparison with GPC, the mean value of SSV has increment from landraces to cultivars (Fig. S2d) and highest values were in the cultivars (Table S1).

GWAS

The filtered SNPs were used to perform GWAS with the CMLM program which greatly reduced false positive rates (Lipka et al. 2012). The GWAS results for PH were presented in Fig. 2a and Fig. S3. Fifty-seven SNPs significantly associated with PH located in 50 different genomic regions in 19 of the 21 chromosomes were observed (Table S2). Six of the 50 regions had two more SNPs with significant association. Nine SNPs around two of the six regions mapped to the same region but more narrow interval of *qPh-6B* on chromosome 6B which was reported in previous study (Zhang et al. 2017) (Table 1). Ten of the SNPs were found in more than two environments. Four of them (a haplotype) were organized as a cluster distributed at physical position 75.0–77.2 Mb segment of chromosome 2A. This haplotype was detected in all the environments and exhibited

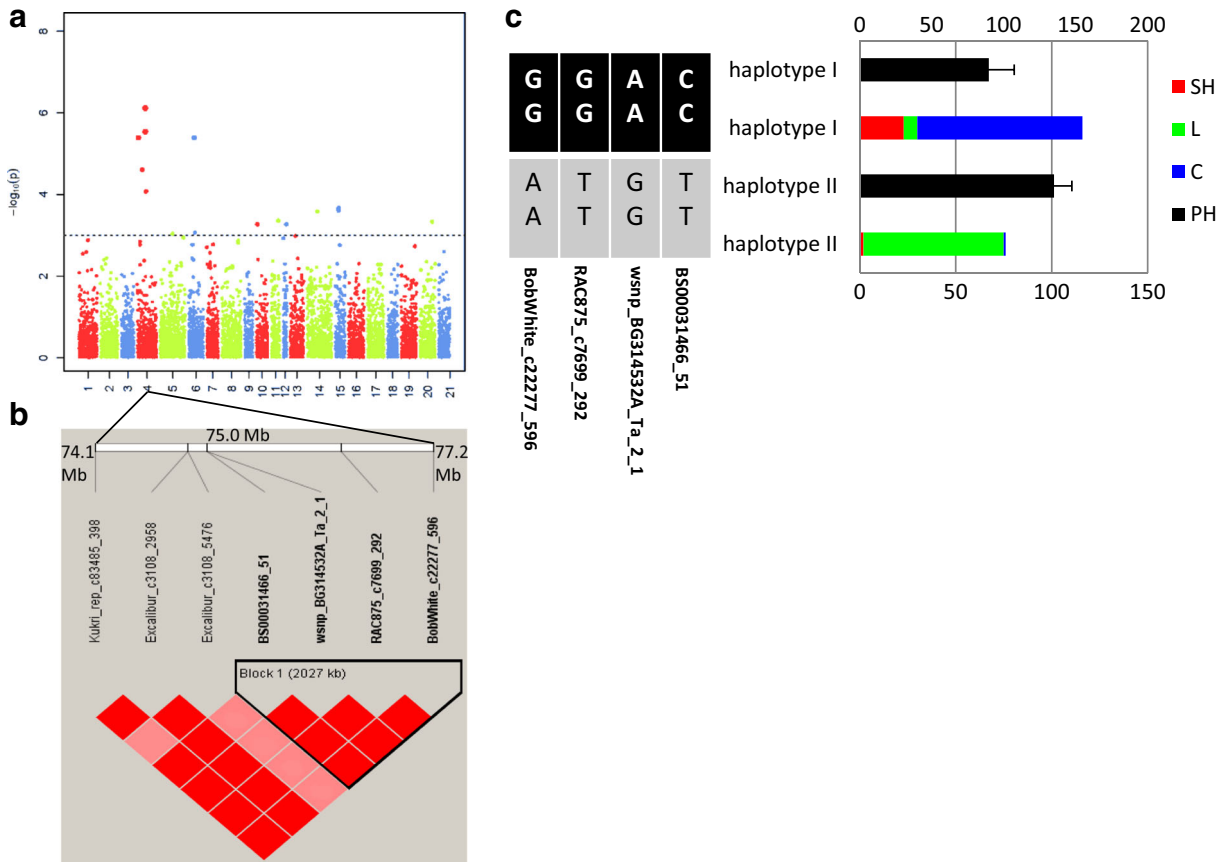


Fig. 2 GWAS results for PH in E2 environment. **a** Manhattan plot for PH in E2 environment. The dashed horizontal line depicted a significant threshold level. **b** A strong linkage disequilibrium region was observed on chromosome 2A from 75.0–77.2 Mb. The four significant SNPs were indicated with bold font. **c**

Haplotypes of the significant SNPs among wheat lines. Stacked bar chart indicated the proportion of each wheat lines for haplotype I and II. Bar plot displayed average PH of each haplotype. *SH* synthetic hexaploid wheat lines, *L* landraces, *C* cultivars, *PH* plant height. Error bars indicated standard error

strongest association with PH (Fig. 2b, Table S2). In addition, the significant association of GWAS results with GPC, TKW, and SSV were detected for 27, 30, and 34 SNPs, respectively (Fig. S3, Table S3–5).

Genome scan and RNA-Seq analysis

Twelve lines used for RNA-Seq analysis were listed on the Supplementary Table S1. The enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis for the differential expression genes was performed to deduce the biological pathway that might be active in the wheat lines. The results showed that the protein processing in endoplasmic reticulum were mostly associated with GPC, and starch/sucrose metabolism with TKW (Fig. S4).

One Mb flanking region of the associated SNPs in wheat genome was scanned to identify genes associated with GPC, TKW, and SSV. To determine whether these found genes were indeed associated with these characters, the differential expression RNA-Seq data was analyzed. The result showed that 102 GPC associated candidate genes, 52 TKW associated candidate genes, and 92 SSV associated candidate genes were actually present and transcribed (Table S6). KEGG analysis showed that two down-regulated expression genes *Traes_2AS_DFDA79E58* and *Traes_7BL_10754D81F* were linked to the pathway involved in proteasome (ko03050), and one up-regulated expression gene *Traes_7AL_1A21E8798* was linked to the pathway involved in protein processing in endoplasmic reticulum (ko04141) (Table S6).

Table 1 SNP clusters associated with PH

SNP	Chromosome	IWGSC wheat sequence survey v2 position	IWGSC RefSeq v1.0 position
BS00031466_51	2A	5,291,608	75,137,557
w SNP_BG314532A_Ta_2_1	2A	5,291,608	75,139,244
RAC875_c7699_292	2A	101,508	76,344,603
BobWhite_c22277_596	2A	3,125,385	77,165,359
BS00090277_51	6B	4,389,506	470,800,994
RAC875_c22078_195	6B	4,348,345	569,044,770
w SNP_Ex_rep_c68452_67273676	6B	4,348,345	569,044,770
Ra_c13513_591	6B	4,280,162	570,954,027
w SNP_JD_rep_c63567_40552622	6B	4,207,479	572,504,679
w SNP_Ku_c11870_19296142	6B	4,247,817	573,324,167
BobWhite_c36416_56	6B	1,070,663	573,689,515
w SNP_Ku_c3185_5949143	6B	1,070,663	573,689,515
w SNP_CAP11_c1185_687769	6B	1,070,663	573,689,726

TaUBP24

PH was an important trait affecting wheat yield (Cui et al. 2011). Its variation was highly associated with a group of polymorphic SNPs detected in a 2.2 Mb (75.0–77.2 Mb) genome region (Fig. 2b). One hundred ninety-two wheat lines were classified into two haplotypes based on the genotyping results. Haplotype I included 23 synthetic hexaploid wheat lines (115.57 ± 14.69 cm), seven landraces (104.94 ± 21.16 cm), and 86 cultivars (81.53 ± 8.07 cm), and haplotype II included two synthetic hexaploid wheat lines (147.53 ± 4.01 cm), 73 landraces (135.32 ± 10.99 cm), and only one cultivar (86.80 cm). The PH mean value of the haplotype I (89.69 ± 17.63 cm) was significantly shorter than the haplotype II (135.00 ± 12.31 cm) (Fig. 2c, Table S7).

Furthermore, the last two of the four significant SNPs belonged to the same gene and located in the fourth intron and the seventh exon of the gene. This gene was homologous to *At4g30890* (*UBP24*, ubiquitin-specific protease 24) designated it as *TaUBP24* (GenBank accession number: MF179095). It contains seven exons and CDS length was 2247 bps (Fig. 3a). The SNP located in the seventh exon was nonsynonymous SNP which could cause amino acid changes from His to Arg (Fig. 3b). The amino acid sequence of TaUBP24 was aligned with reported sequences from *Brachypodium distachyon*, *Oryza sativa* spp. japonica, and *Arabidopsis thaliana*. All the sequences were checked for the presence of the important active sites (Asn-Cys-

His-Asp) (Hu et al. 2002) which were essential for the enzyme activity (Fig. 3b).

For phylogenetic analysis, UBP24 amino acid sequences from *Triticum aestivum*, *Arabidopsis thaliana*, *Hordeum vulgare* subsp. *vulgare*, *Brachypodium distachyon*, *Oryza sativa* spp. japonica, *Setaria italica*, *Zea mays*, *Sorghum bicolor*, and *Glycine max* were analyzed. The homologous sequence of *Homo sapiens* was used as an outgroup control. The neighbor-joining program with the parameter selection of *P* distance model, pairwise deletion, and 1000 bootstrap was conducted. As a result, monocots and dicots were well separated into two clades (Fig. 4).

For marker-assisted selection, two pairs of complementary type primers were developed to distinguish the two haplotypes (Table S7). The nonsynonymous SNP located in the seventh exon of *TaUBP24* was employed (Fig. 2c). The PCR product could be amplified by TaUBP24-L primer pair in the haplotype I, which could not be detected in haplotype II. TaUBP24-H primer pair could amplify in the haplotype II and not in haplotype I (Fig. S5). Thus, the molecular markers based on PCR reliably discriminated the two haplotypes.

Discussion

The association study approach was firstly proposed by Risch and Merikangas (1996). The application of GWAS in plant for the first time was analyzed the growth habit of

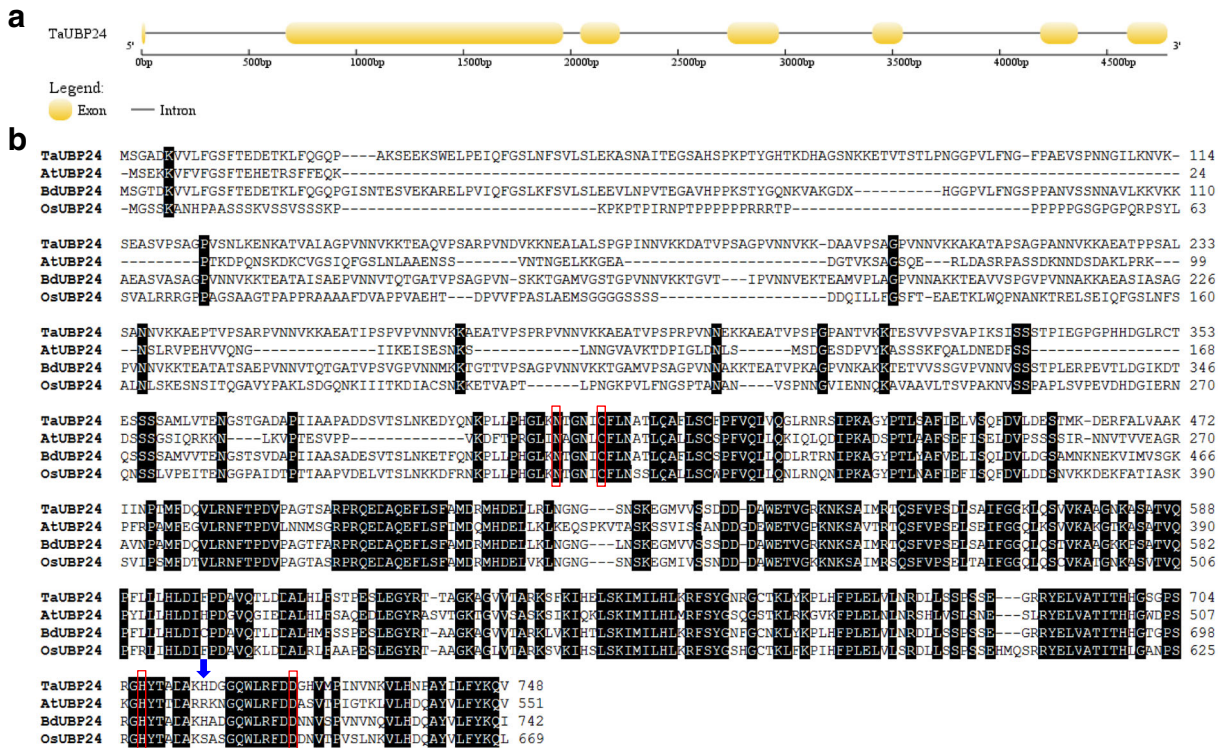


Fig. 3 Analysis of TaUBP24. **a** Gene structure of TaUBP24. **b** Comparison of the amino acid sequences of UBP24 from *Triticum aestivum* (TaUBP24), *Brachypodium distachyon* (BdUBP24), *Oryza sativa* spp. *japonica*, (OsUBP24), and

Arabidopsis thaliana (AtUBP24). The residues in the red solid line frames were the active sites required for catalytic core domain of all UBP24. The arrow indicated the nonsynonymous SNP locus

sea beet (Hansen et al. 2001). With the development of high-throughput sequencing technologies, GWAS has been widely used for dissecting the genetic basis of

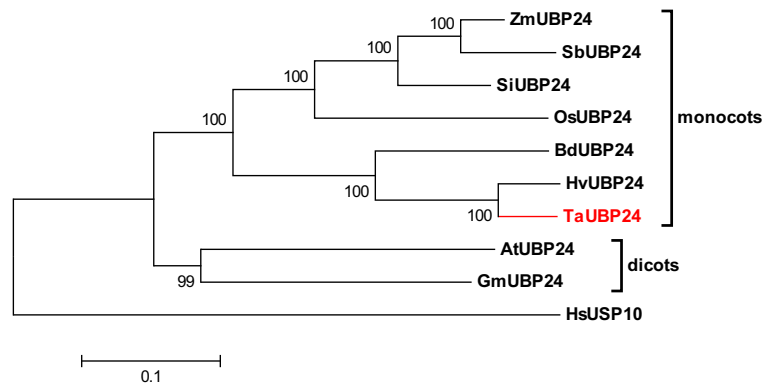


Fig. 4 Phylogenetic analysis of UBP24. Two clades were present on the basis of monocots and dicots. TaUBP24 was indicated in red font. *Triticum aestivum* (TaUBP24), *Arabidopsis thaliana* (AtUBP24, BAH19813), *Hordeum vulgare* subsp. *vulgare*, (HvUBP24, BAJ92714), *Brachypodium distachyon* (BdUBP24,

XP_010234436), *Oryza sativa* spp. *japonica*, (OsUBP24, XP_015646906), *Setaria italica* (SiUBP24, XP_004958591), *Zea mays* (ZmUBP24, XP_020396334), *Sorghum bicolor* (SbUBP24, OQU90420), *Glycine max* (GmUBP24, XP_003539934), and *Homo sapiens* (HsUSP10, NP_001259004)

QTL mapping, GWAS overcame the limitation of biparental populations and enhanced the genome resolution often to the gene level, reduced the cost and saved time for its power to detect whole genomic variations from large sample size (Brachi et al. 2011).

GPC and SSV were two important traits affecting grain quality (Ormoli et al. 2015; Prasad et al. 2003). In this study, GWAS and RNA-Seq were carried out at the same time and several genes associated with GPC, TKW, and SSV were identified. Among these genes, Traes_2AS_DFDA79E58 and Traes_7BL_10754D81F linked to the pathway of proteasome possibly taking part in the process of proteasome-mediated protein degradation, which were down-regulated expression and consequently affected grain quality (Guo et al. 2015). Traes_7AL_1A21E8798, an up-regulated expression gene, which was involved in the pathway of protein processing in endoplasmic reticulum might participate in protein accumulation within the endoplasmic reticulum directly resulted in GPC increased (Tosi et al. 2009).

PH was a crucial trait affecting grain yield (Cui et al. 2011). Since the “Green Revolution” period, dwarf genes of *Rht-B1b* and *Rht-D1b* were introduced into crops to resist lodging, which resulted in a rapid increase of grain yield (Gale et al. 1975; Hedden 2003; Li et al. 2012b). There were many other *Rht* (reduced height) genes that have been detected. For example, *Rht_NM9*, *Rht4*, *Rht5*, *Rht8*, *Rht9*, *Rht12*, *Rht13*, *Rht14*, *Rht16*, *Rht18*, *Rht22*, and *Rht23* (Chen et al. 2015; Ellis et al. 2007; Ellis et al. 2005; Haque et al. 2011; Lu et al. 2015; Peng et al. 2011). In this study, two groups of strong GWAS signals were detected. One was mapped to the previously identified plant height QTL *qPh-6B* (Zhang et al. 2017), and another one located on chromosome 2A encompassing the gene *TaUBP24*.

Abscisic acid (ABA) and gibberellic acid (GA) were notable mutual antagonism phytohormones which have been proved to affect PH (Yaish et al. 2010). Dwarfing gene *OsAP2-39* has been confirmed to increase endogenous ABA level through up-regulating ABA biosynthetic gene *OsNCED-1* which could make the EUI protein up-regulate expression to deactivate gibberellins in rice (Yaish et al. 2010). In *Arabidopsis thaliana*, the *UBP24* mutant was hypersensitive to ABA in both post-germinative stage and seedling growth. *UBP24* had deubiquitinating enzyme activity which was essential for *UBP24* function. *UBP24* genetically located upstream of *ABI2*, and the phosphatase activity of protein phosphatase 2C (PP2Cs including *ABI1* and *ABI2*) was

lower in the *UBP24* mutant than that in the wild type in the presence of ABA. These results supported the important regulatory role for the *UBP24* that negatively regulate ABA signaling in plant (Zhao et al. 2016). To sum up, the wheat homologs of *Arabidopsis thaliana* *UBP24*, *TaUBP24* might indirectly up-regulate GA expression in the pathway and consequently affecting PH.

In our study, two significant SNPs located in *TaUBP24*, together with the other two SNPs were classified into two haplotypes in 192 wheat lines. The PH was significantly different between haplotype I and II. Also, the synthetic hexaploid wheat lines were significantly higher than the cultivars in haplotype I and II (Table S8). The possible reason might be that synthetic hexaploid wheat was obtained by crossing of durum wheat and *Aegilops tauschii* and did not subject to artificial selection (Zegeye et al. 2014). PH was complex quantitative trait controlled by multiple genes (Zhang et al. 2017). The effect of *TaUBP24* gene on the PH was highlighted in the cultivars by selection breeding.

In conclusion, several candidate genes associated with GPC, TKW, and SSV were found by using GWAS and RNA-Seq. *TaUBP24*, encoding an ubiquitin-specific protease, was first identified to be associated with PH. The functional markers of *TaUBP24* could be directly used for marker-assisted selection to improve wheat quality and yield.

Acknowledgments This work was supported by grants from the National Basic Research Program of China (973 Program, 2014CB138104-2), The National Key Research and Development Program of China (Grant No. 2016YFD0102000), and the National Key Project of Transgenic Biologic Varieties Breeding of China (Grant No. 2016ZX08009003-004).

Compliance with ethical standards

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

References

- AACC International (1999) Approved methods of analysis, 11th Ed. Method 56-61A. Sedimentation test for wheat. Approved November 3. AACC International, St. Paul. <https://doi.org/10.1094/AACCIIntMethod-56-61.02>
- Ain Q, Rasheed A, Anwar A, Mahmood T, Imtiaz M, Mahmood T, Xia X, He Z, Quraishi UM (2015) Genome-wide association for grain yield under rainfed conditions in historical wheat cultivars from Pakistan. *Front Plant Sci* 6:743. <https://doi.org/10.3389/fpls.2015.00743>

- Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664. <https://doi.org/10.1101/gr.094052.109>
- Brachi B, Morris GP, Borevitz JO (2011) Genome-wide association studies in plants: the missing heritability is in the field. *Genome Biol* 12(10):232. <https://doi.org/10.1186/Gb-2011-12-10-232>
- Breseghele F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172(2):1165–1177. <https://doi.org/10.1534/genetics.105.044586>
- Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* 4(1):7. <https://doi.org/10.1186/s13742-015-0047-8>
- Chen HD, Xie W, He H, Yu H, Chen W, Li J, Yu R, Yao Y, Zhang W, He Y, Tang X, Zhou F, Deng XW, Zhang Q (2014) A high-density SNP genotyping array for rice biology and molecular breeding. *Mol Plant* 7(3):541–553. <https://doi.org/10.1093/mp/sst135>
- Chen S, Gao R, Wang H, Wen M, Xiao J, Bian N, Zhang R, Hu W, Cheng S, Bie T, Wang X (2015) Characterization of a novel reduced height gene (*Rht23*) regulating panicle morphology and plant architecture in bread wheat. *Euphytica* 203(3):583–594. <https://doi.org/10.1007/s10681-014-1275-1>
- Cook JP, McMullen MD, Holland JB, Tian F, Bradbury P, Ross-Ibarra J, Buckler ES, Flint-Garcia SA (2012) Genetic architecture of maize kernel composition in the nested association mapping and inbred association panels. *Plant Physiol* 158(2):824–834. <https://doi.org/10.1104/pp.111.185033>
- Cui F, Li J, Ding A, Zhao C, Wang L, Wang X, Li S, Bao Y, Li X, Feng D, Kong L, Wang H (2011) Conditional QTL mapping for plant height with respect to the length of the spike and internode in two mapping populations of wheat. *Theor Appl Genet* 122(8):1517–1536. <https://doi.org/10.1007/s00122-011-1551-6>
- Curtis T, Halford NG (2014) Food security: the challenge of increasing wheat yield and the importance of not compromising food safety. *Ann Appl Biol* 164(3):354–372. <https://doi.org/10.1111/aab.12108>
- Distelfeld A, Korol A, Dubcovsky J, Uauy C, Blake T, Fahima T (2008) Colinearity between the barley grain protein content (GPC) QTL on chromosome arm 6HS and the wheat *Gpc-B1* region. *Mol Breeding* 22(1):25–38. <https://doi.org/10.1007/s11032-007-9153-3>
- Echeverry-Solarte M, Kumar A, Kianian S, Simsek S, Alamri MS, Mantovani EE, McClean PE, Deckard EL, Elias E, Schatz B, Xu SS, Mergoum M (2015) New QTL alleles for quality-related traits in spring wheat revealed by RIL population derived from supernumerary×non-supernumerary spikelet genotypes. *Theor Appl Genet* 128(5):893–912. <https://doi.org/10.1007/s00122-015-2478-0>
- Ellis MH, Rebetzke GJ, Azaña F, Richards RA, Spielmeier W (2005) Molecular mapping of gibberellin-responsive dwarfing genes in bread wheat. *Theor Appl Genet* 111(3):423–430. <https://doi.org/10.1007/s00122-005-2008-6>
- Ellis MH, Bonnett DG, Rebetzke GJ (2007) A 192bp allele at the *Xgwm261* locus is not always associated with the *Rht8* dwarfing gene in wheat (*Triticum aestivum* L.). *Euphytica* 157(1-2):209–214. <https://doi.org/10.1007/s10681-007-9413-7>
- Gale MD, Law CN, Worland AJ (1975) Chromosomal location of a major dwarfing gene from Norin 10 in new British semi-dwarf wheats. *Heredity* 35(3):417–421. <https://doi.org/10.1038/hdy.1975.112>
- Gao L, Turner MK, Chao SM, Kolmer J, Anderson JA (2016) Genome wide association study of seedling and adult plant leaf rust resistance in elite spring wheat breeding lines. *PLoS One* 11(2):e0148671. <https://doi.org/10.1371/journal.pone.0148671>
- Groos C, Robert N, Bervas E, Charmet G (2003) Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. *Theor Appl Genet* 106(6):1032–1040. <https://doi.org/10.1007/s00122-002-1111-1>
- Guo W, Zhang J, Zhang N, Xin M, Peng H, Hu Z, Ni Z, Du J (2015) The wheat NAC transcription factor TaNAC2L is regulated at the transcriptional and post-translational levels and promotes heat stress tolerance in transgenic Arabidopsis. *PLoS One* 10(8):e0135667. <https://doi.org/10.1371/journal.pone.0135667>
- Hansen M, Kraft T, Ganestam S, Sall T, Nilsson NO (2001) Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. *Genet Res* 77(1):61–66. <https://doi.org/10.1017/S0016672300004857>
- Hao DR, Chao MN, Yin ZT, Yu DY (2012) Genome-wide association analysis detecting significant single nucleotide polymorphisms for chlorophyll and chlorophyll fluorescence parameters in soybean (*Glycine max*) landraces. *Euphytica* 186(3):919–931. <https://doi.org/10.1007/s10681-012-0697-x>
- Haque MA, Martinek P, Watanabe N, Kuboyama T (2011) Genetic mapping of gibberellic acid-sensitive genes for semi-dwarfism in durum wheat. *Cereal Res Commun* 39(2):171–178. <https://doi.org/10.1556/Crc.39.2011.2.1>
- Hedden P (2003) The genes of the green revolution. *Trends Genet* 19(1):5–9. [https://doi.org/10.1016/S0168-9525\(02\)00009-4](https://doi.org/10.1016/S0168-9525(02)00009-4)
- Hu M, Li P, Li M, Li W, Yao T, Wu JW, Gu W, Cohen RE, Shi Y (2002) Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. *Cell* 111(7):1041–1054. [https://doi.org/10.1016/S0092-8674\(02\)01199-6](https://doi.org/10.1016/S0092-8674(02)01199-6)
- Hu B, Jin JP, Guo AY, Zhang H, Luo JC, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31(8):1296–1297. <https://doi.org/10.1093/bioinformatics/btu817>
- Huang X, Wei X, Sang T, Zhao QA, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42(11):961–U976. <https://doi.org/10.1038/ng.695>
- Kollers S, Rodemann B, Ling J, Korzun V, Ebmeyer E, Argillier O, Hinze M, Plieske J, Kulosa D, Gana MW, Röder MS (2013) Whole genome association mapping of fusarium head blight resistance in European winter wheat (*Triticum aestivum* L.). *PLoS One* 8(2):e57500. <https://doi.org/10.1371/journal.pone.0057500.g001>
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23(21):2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>

- Li YL, Zhou RH, Wang J, Liao XZ, Branlard G, Jia JZ (2012a) Novel and favorable QTL allele clusters for end-use quality revealed by introgression lines derived from synthetic wheat. *Mol Breeding* 29(3):627–643. <https://doi.org/10.1007/s11032-011-9578-6>
- Li Y, Xiao J, Wu J, Duan J, Liu Y, Ye X, Zhang X, Guo X, Gu Y, Zhang L, Jia J, Kong X (2012b) A tandem segmental duplication (TSD) in green revolution gene *Rht-D1b* region underlies plant height variation. *New Phytol* 196(1):282–291. <https://doi.org/10.1111/j.1469-8137.2012.04243.x>
- Lipka AE, Tian F, Wang QS, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang ZW (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28(18):2397–2399. <https://doi.org/10.1093/bioinformatics/bts444>
- Lopes MS, Reynolds MP, McIntyre CL, Mathews KL, Jalal Kamali MR, Mossad M, Feltaous Y, Tahir ISA, Chatrath R, Ogbonnaya F, Baum M (2013) QTL for yield and associated traits in the Seri/Babax population grown across several environments in Mexico, in the West Asia, North Africa, and South Asia regions. *Theor Appl Genet* 126(4):971–984. <https://doi.org/10.1007/s00122-012-2030-4>
- Lu Y, Xing L, Xing S, Hu P, Cui C, Zhang M, Xiao J, Wang H, Zhang R, Wang X, Chen P, Cao A (2015) Characterization of a putative new semi-dominant reduced height gene, *Rht_NM9*, in wheat (*Triticum aestivum* L.) *J Genet Genomics* 42(12):685–698. <https://doi.org/10.1016/j.jgg.2015.08.007>
- Maccaferri M, Zhang JL, Bulli P, Abate Z, Chao SM, Cantu D, Bossolini E, Chen XM, Pumphrey M, Dubcovsky J (2015) A genome-wide association study of resistance to stripe rust (*Puccinia striiformis* f. sp. *tritici*) in a worldwide collection of Hexaploid spring wheat (*Triticum aestivum* L.) G3-Genes *Genom Genet* 5(3):449–465. <https://doi.org/10.1534/g3.114.014563>
- Marcotuli I, Houston K, Waugh R, Fincher GB, Burton RA, Blanco A, Gadaleta A (2015) Genome wide association mapping for arabinoxylan content in a collection of tetraploid wheats. *PLoS One* 10(7):e0132787. <https://doi.org/10.1371/journal.pone.0132787>
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular-weight plant DNA. *Nucleic Acids Res* 8(19):4321–4325. <https://doi.org/10.1093/nar/8.19.4321>
- Naruoka Y, Garland-Campbell KA, Carter AH (2015) Genome-wide association mapping for stripe rust (*Puccinia striiformis* F. sp. *tritici*) in US Pacific Northwest winter wheat (*Triticum aestivum* L.) *Theor Appl Genet* 128(6):1083–1101. <https://doi.org/10.1007/s00122-015-2492-2>
- Neumann K, Kobiljski B, Denčić S, Varshney RK, Börner A (2010) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.) *Mol Breeding* 27(1):37–58. <https://doi.org/10.1007/s11032-010-9411-7>
- Nigro D, Blanco A, Anderson OD, Gadaleta A (2014) Characterization of *Ferredoxin-Dependent Glutamine-Oxoglutarate Amidotransferase (Fd-GOGAT)* genes and their relationship with grain protein content QTL in wheat. *PLoS One* 9(8):e0103869. <https://doi.org/10.1371/journal.pone.0103869>
- Ormolli L, Costa C, Negri S, Perenzin M, Vaccino P (2015) Diversity trends in bread wheat in Italy during the 20th century assessed by traditional and multivariate approaches. *Sci Rep-Uk* 5(1):8574. <https://doi.org/10.1038/srep08574>
- Peng ZS, Li X, Yang ZJ, Liao ML (2011) A new reduced height gene found in the tetraploid semi-dwarf wheat landrace Aiganfanmai. *Genet Mol Res* 10(4):2349–2357. <https://doi.org/10.4238/2011.October.5.5>
- Prasad M, Kumar N, Kulwal PL, Roder MS, Balyan HS, Dhaliwal HS, Gupta PK (2003) QTL analysis for grain protein content using SSR markers and validation studies using NILs in bread wheat. *Theor Appl Genet* 106(4):659–667. <https://doi.org/10.1007/s00122-002-1114y>
- Rasheed A, Xia XC, Ogbonnaya F, Mahmood T, Zhang ZW, Mujeeb-Kazi A, He ZH (2014) Genome-wide association for grain morphology in synthetic hexaploid wheats using digital imaging analysis. *BMC Plant Biol* 14(1):128. <https://doi.org/10.1186/1471-2229-14-128>
- Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. *Science* 273(5281):1516–1517. <https://doi.org/10.1126/science.273.5281.1516>
- Sukumaran S, Dreisigacker S, Lopes M, Chavez P, Reynolds MP (2015) Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor Appl Genet* 128(2):353–363. <https://doi.org/10.1007/s00122-014-2435-3>
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30(12):2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Terasawa Y, Ito M, Tabiki T, Nagasawa K, Hatta K, Nishio Z (2016) Mapping of a major QTL associated with protein content on chromosome 2B in hard red winter wheat (*Triticum aestivum* L.) *Breeding Sci* 66(4):471–480. <https://doi.org/10.1270/jsbbs.16026>
- Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler E (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet* 43(2):159–U113. <https://doi.org/10.1038/Ng.746>
- Tosi P, Parker M, Gritsch CS, Carzaniga R, Martin B, Shewry PR (2009) Trafficking of storage proteins in developing grain of wheat. *J Exp Bot* 60(3):979–991. <https://doi.org/10.1093/jxb/ern346>
- USDA (2017) World agricultural supply and demand estimates. <https://www.usda.gov/oce/commodity/wasde/latest.pdf>
- Wang S, 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, Ganai M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol J* 12(6):787–796. <https://doi.org/10.1111/pbi.12183>
- Yaish MW, El-kereamy A, Zhu T, Beatty PH, Good AG, Bi YM, Rothstein SJ (2010) The APETALA-2-like transcription factor OsAP2-39 controls key interactions between abscisic acid and gibberellin in rice. *PLoS Genet* 6(9):e1001098. <https://doi.org/10.1371/journal.pgen.1001098>

- Yu LX, Lorenz A, Rutkoski J, Singh RP, Bhavani S, Huerta-Espino J, Sorrells ME (2011) Association mapping and gene-gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. *Theor Appl Genet* 123(8):1257–1268. <https://doi.org/10.1007/s00122-011-1664-y>
- Zanke C, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Argillier O, Stiewe G, Hinze M, Beier S, Ganal MW, Röder MS (2014a) Genetic architecture of main effect QTL for heading date in European winter wheat. *Front Plant Sci* 5:217. <https://doi.org/10.3389/fpls.2014.00217>
- Zanke CD, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Neumann K, Ganal MW, Röder MS (2014b) Whole genome association mapping of plant height in winter wheat (*Triticum aestivum* L.) PLoS One 9(11):e113287. <https://doi.org/10.1371/journal.pone.0113287>
- Zanke CD, Ling J, Plieske J, Kollers S, Ebmeyer E, Korzun V, Argillier O, Stiewe G, Hinze M, Neumann F, Eichhorn A, Polley A, Jaenecke C, Ganal MW, Röder MS (2015) Analysis of main effect QTL for thousand grain weight in European winter wheat (*Triticum aestivum* L.) by genome-wide association mapping. *Front. Plant Sci* 6:644. <https://doi.org/10.3389/fpls.2015.0064>
- Zegeye H, Rasheed A, Makdis F, Badebo A, Ogonnaya FC (2014) Genome-wide association mapping for seedling and adult plant resistance to stripe rust in synthetic hexaploid wheat. *PLoS One* 9(8):e105593. <https://doi.org/10.1371/journal.pone.0105593>
- Zhang N, Fan X, Cui F, Zhao C, Zhang W, Zhao X, Yang L, Pan R, Chen M, Han J, Ji J, Liu D, Zhao Z, Tong Y, Zhang A, Wang T, Li J (2017) Characterization of the temporal and spatial expression of wheat (*Triticum aestivum* L.) plant height at the QTL level and their influence on yield-related traits. *Theor Appl Genet* 123(6):1257–1268. <https://doi.org/10.1007/s00122-017-2884-6>
- Zhao J, Zhou H, Zhang M, Gao Y, Li L, Gao Y, Li M, Yang Y, Guo Y, Li X (2016) Ubiquitin-specific protease 24 negatively regulates abscisic acid signalling in *Arabidopsis thaliana*. *Plant Cell Environ* 39(2):427–440. <https://doi.org/10.1111/pce.12628>
- Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, Yu Y, Shu L, Zhao Y, Ma Y, Fang C, Shen Y, Liu T, Li C, Li Q, Wu M, Wang M, Wu Y, Dong Y, Wan W, Wang X, Ding Z, Gao Y, Xiang H, Zhu B, Lee SH, Wang W, Tian Z (2015) Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. *Nat Biotechnol* 33(4):408–414. <https://doi.org/10.1038/nbt.3096>