Euphytica (2019) 215:201
https://doi.org/10.1007/s10681-019-2517-z

Genome-wide association study of total starch and its components in common wheat

Xiaoying Chen · Wenqi Fang · Mengqi Ji · Shuai Xu · Yanxin Jiang · Shu Song · Guangfeng Chen · Jichun Tian · Zhiying Deng

Received: 10 April 2019 / Accepted: 25 October 2019 / Published online: 14 November 2019

Abstract Wheat starch is closely related to yield and quality of wheat-based final products. This genome-wide association study focused on total starch content and its components using single nucleotide polymorphisms (SNPs) in a panel of 205 elite winter wheat accessions. Twenty-nine significant marker-trait associations (MTAs) were detected for total starch (TSC), amylose (AMS) and amylopectin (AMP) contents under four environmental regimes. Nine MTAs were detected for two traits, and eleven MTAs were found for all three traits. These SNPs were distributed across seven chromosomes, explaining 11.26–23.83% of the phenotypic variance (PVE). Furthermore, eighteen MTAs related to the ratio of AMS and AMP under four environmental conditions explained 5.92–17.2% of the phenotypic variance. One important multi-trait MTA on chromosome 3A at the 93 cM position was found for thousand kernel weight, AMP, AMS and TSC. A set of elite alleles was identified, including allele *Kukri_c5615_1214-A*, which increased AMS and AMP by 5.80% and 15.32%, respectively. The allele *Excalibur_c16376_351-T* increased AMP by 4.43%, and the T allele of marker *BS00022255_51* and T allele of marker *D_contig25392_201* had the most significant effects, increasing the ratio of AMS to AMP by 3.77%. Fourteen candidate genes associated with significant markers were identified. For example, *TRIAE_CS42_2AL_TGACv1_093900_AA0288950* participated in carbohydrate metabolism and hydrolysed O-glycosyl compounds. The lines that carry these elite alleles could be used as genetic stock for breeding to improve these traits. These results could lay the foundation for molecular marker-assisted selection in improving wheat quality.

Keywords Amylose · Amylopectin · Genome-wide association study · Starch content

Xiaoying Chen and Wenqi Fang contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10681-019-2517-z) contains supplementary material, which is available to authorized users.

X. Chen · W. Fang · M. Ji · S. Xu · Y. Jiang · S. Song · J. Tian · Z. Deng (✉)
State Key Laboratory of Crop Biology, Key Laboratory of Crop Biology of Shandong Province, Group of Wheat Quality and Molecular Breeding, College of Agronomy, Shandong Agricultural University,
Tai'an 271000, Shandong, People's Republic of China
e-mail: zydeng@sdau.edu.cn

G. Chen
College of Ecology and Garden Architecture, Dezhou University, Dezhou 253023, China

Introduction

In the mature common wheat kernel, starch accounts for approximately three-quarters of the grain composition and contains 20–30% amylose (AMS) and 70–80% amylopectin (AMP). Starch also accounts for approximately 65 to 70% of the dry grain weight (Li et al. 1999). Endosperm starch content not only influences the grain weight but also grain processing and end-use quality (Hurkman et al. 2003). In fact, not only starch content but its components, properties and functions also affect food quality, such as appearance, flavour, texture, and nutritional value (Martin and Smith 1995; Liu et al. 2003; Song et al. 2008). Starch components, especially the ratio of the components, are very important for processing and end-use quality. The varietal differences in the AMP structure are predominantly due to chain length variation and play a critical role in determining the physicochemical properties of starch in wheat endosperm. AMS content is an important indicator of wheat grain quality. Previous studies showed that AMS was significantly correlated with the quality of noodles, bread and steamed buns. There was a negative correlation between AMS and noodle quality, that is, a high AMS content is associated with a reduced overall noodle score (Wang et al. 1998). In contrast, in sensory evaluations, as the amount of AMP increased, the scores for cohesiveness, springiness, and acceptability of cooked noodles also increased. The proper AMS: AMP ratio improved the freeze–thaw stability and the sensory acceptability of wheat flour dough and noodles (Cho et al. 2007). Steamed bread made from flour with a low AMS content was bulky and had good eating quality (Liu et al. 2003), and a significant positive correlation was found between AMP content in flour and the quality of bread (Zhou 2012). AMP forms a network structure in dough, resulting in a porous material that produces more water vapour during the expansion process, improves bread volume, texture and palatability. The characteristics and functions of AMP are exactly opposite to those of AMS, and the smaller the ratio of AMS: AMP, the more favourable the processing quality of gluten-based food.

Some key enzymes are involved in the biosynthesis of AMS and AMP. AMS is mainly controlled by granule bound starch synthase I (GBSS I), which is the waxy protein whose encoded genes are located on chromosomes 4AL (*Wx-B1*), 7AS (*Wx-A1*) and 7DS (*Wx-D1*) (Nakamura et al. 1993; Martin et al. 2004). A lack of waxy genes affects the AMS content and further influences starch quality. AMP synthesis involves several key enzymes, such as soluble starch synthase (SSS, involving SSI, SSIIa, SSIII), debranching enzymes (DBE, containing DBEI), and branching enzymes (BE, involving BEI, BEIIa, BEIIb) (Li et al. 2003; Nakamura et al. 2017; Crofts et al. 2017). The expression of these enzymes affects the biosynthesis of AMP. The physical and chemical properties of starch are directly affected by the AMS: AMP ratio. Therefore, expression of these enzymes influences the contents and ratio of AMS to AMP and ultimately affects processing quality (Caballero et al. 2008). Traditional quantitative trait loci (QTL) mapping and genome-wide association studies (GWAS) have been used as the main methods for dissecting complex traits (Risch and Merikangas 1996). Many QTL mapping studies have been performed to explore and characterize the genetic basis of accumulation of wheat starch and its components (Araki et al. 1999; Batey et al. 2002; Deng et al. 2015, 2018). QTLs were mainly distributed on chromosomes 1A, 1D, 2A, 2D, 3B, 4A, 5A, 7A, 7B and 7D in various populations (McCartney et al. 2006; Sun et al. 2008; Tian et al. 2015; Deng et al. 2017). Although traditional QTL mapping has been very successful for genetic analysis when exploring genetic variation, its relevance is limited to the genomes of only two parents.

GWAS, as a complement to QTL mapping, is the most cost-effective way to use existing germplasm (such as landraces, elite cultivars, and advanced breeding lines) for genetic mapping (Newell et al. 2012; Bradbury et al. 2011; Bandillo et al. 2015). With the rapid development of high-density marker-assisted genotyping techniques and next-generation sequencing (NGS), GWAS has become a widely used method for identifying the genes responsible for the quantitative variation of complex traits (Zhu et al. 2008). This strategy has been successfully used for agronomic traits in rice, maize, barley, common wheat, durum wheat and other crops (Huang et al. 2011; Cockram et al. 2010; Yu and Buckler 2006; Chen et al. 2015; Li et al. 2017a; Ovenden et al. 2017; Liu et al. 2016; Shu et al. 2012).

To improve the resolution of association maps and to cover the entire genome with sufficient resolution, a large number of molecular markers are needed for

GWAS (Sajjad et al. 2012). Single nucleotide polymorphism (SNP) markers representing third-generation molecular markers are abundant and evenly distributed across genomes, satisfying the large sample and high-density marker requirements of GWAS (Gupta et al. 2008). At present, GWAS using SNPs has been widely used to illuminate the genetics of many animals and plants, such as humans (Mick et al. 2011), rice (Huang et al. 2011), and maize (Wilson et al. 2004). However, GWAS in wheat continues to be a challenge due its complex genomic architecture and incomplete genome sequence (Sukumaran and Yu 2014). Recently, several SNP-based technologies, such as genotyping chips, have become available. For example, the 90 K Illumina iSelect array (Wang et al. 2014) is commonly used in wheat genetics and breeding research, including genetic mapping and association analysis of important agronomic traits.

Compared to traditional QTL mapping, association mapping studies has advantages especially in increased QTL resolution and allele coverage. Although previous researchers have dissected the genetic basis for the accumulation of wheat starch and its components using QTL mapping, GWAS was rarely found in their analysis. Therefore, the present study used GWAS to dissect total starch and its components, with 24,355 SNPs genotyped using the 90 K Illumina iSelect array in a population of diverse winter wheat varieties. The objectives of this study were to identify markers and candidate genes for loci associated with these traits in order to improve wheat starch quality by breeding.

Materials and methods

Plant material and growth conditions

The association mapping panel of 205 wheat genotypes for GWAS comprised 77 released cultivars, 55 founder parents, and 73 breeding lines (Table S1) from 10 provinces that represent the major winter wheat production regions in China. Two lines from Mexico and France were included as additional founder parents.

The panel was grown in the 2013–2014 and 2014–2015 cropping seasons in experimental fields at Shandong Agricultural University, Tai'an (116°36′E, 36°57′N) and Dezhou Institute of Agricultural Sciences (116°29′E, 37°45′N). The experimental fields were arranged in randomized block design, with two replicates for each environment. All lines were grown in 2 m plots with 3 rows spaced 25 cm apart, and 70 seeds were evenly spaced in each row. Field management followed local procedures. No serious pest damage or lodging problems occurred during the trials.

Measurement of starch components

Starch, AMS and AMP contents were measured by the double-wave method (Jin et al. 2009) with modifications. The main wavelength for determining AMS content was 471 nm, and the comparison wavelength was 632 nm. The main wavelength for determining AMP content was 553 nm, and the comparison wavelength was 740.3 nm. The AMS and AMP contents in each sample were determined according to the extracted dilution factor relationship, and the total starch content (TSC) was taken as the sum of the AMS and AMP contents.

Analysis of phenotypic data

Analysis of variance (ANOVA) and correlations among phenotypic traits were carried out using SPSS version 17.0 (SPSS Inc., Chicago, IL, USA). Heritability (h^2) was calculated as $h_B^2 = \sigma_g^2/(\sigma_g^2 + \sigma_{ge}^2/r + \sigma_e^2/re)$, where σ_g^2, σ_{ge}^2, and σ_e^2 were estimates of genotype, genotype × environment and residual error variances, respectively. Estimates of σ_g^2, σ_{ge}^2, and σ_e^2 were obtained from the ANOVA, which was performed using the PROC GLM procedure in SAS 8.0 (SAS Institute Inc., Cary, NC, USA).

SNP markers and genotyping

SNP genotyping was performed at the University of California, Davis Genome Center. An Illumina iScan Reader was used to carry out the genotyping assays (Chen et al. 2016). The genetic diversity data were reported previously (Chen et al. 2016, 2017).

DNA extraction and a composite genetic map

DNA was extracted from the young leaf tissues of each variety following to the method recommended by Triticarte Pty. Ltd. (http://www.triticarte.com.au). Samples were genotyped using the 90 K iSelect wheat

chip, which consists of 81,587 SNP loci distributed across all 21 wheat chromosomes.

The total length of the map was 3674.16 cM, with a mean genetic distance of 0.15 cM between markers. Chromosome 1B contained the most markers (n = 2390), followed by 5B (n = 2187), whereas chromosome 4D had the fewest loci (n = 78). Among the A, B and D genomes, the B genome contained the largest number of loci (n = 12,321) and a total length of 1150.47 cM, followed by the A genome (n = 9523) at 1252.51 cM, and the D genome (n = 2511) at 1271.18 cM (Chen et al. 2017).

Population structure

Population structure analysis was performed on genotypic data obtained from unlinked SNP markers in the 205 winter wheat accessions using NJ cluster analysis in STRUCTURE v 2.2 (Chen et al. 2017).

Genome-wide association analysis

Significant marker-trait associations (MTAs) were identified using a mixed linear model (MLM) in TASSEL 3.0. Decisions on whether a QTL was associated with a marker was determined by P value. R^2 values were used as estimates of the magnitude of MTA effects. SNPs with corrected P values ≤ 0.01 were considered to be significantly associated with phenotypic traits.

Identification of candidate genes

To identify the position of important MTA loci in the physical map and to identify possible candidate genes, a BLAST search was performed on the International Wheat Genome Sequencing Consortium database (IWGSC; http://www.wheatgenome.org/, accessed 27th April, 2018) using the sequences of significant SNP markers identified by GWAS. When a SNP marker sequence from the IWGSC was 100% identical to any wheat contig, the sequence was extended 5 kb using the IWGSC BLAST results. The extended sequence was used to run BLAST searches on the National Center for Biotechnology Information (NCBI) database (http://www.ncbi.nlm.nih.gov, 27th April, 2018) and Ensembl Plants (http://plants.ensembl.org/Triticum_aestivum/Tools/BLASt, 27th April, 2018) to confirm possible candidate genes and putative functions.

Results

Population structure

When ΔK values were plotted against hypothetical subgroups the highest ΔK was observed at K = 4, indicating the likelihood of four subgroups in the association panel. Using the maximum membership probability in STRUCTURE, the 205 accessions were segregated into four subpopulations: subgroup 1 (43 accessions), subgroup 2 (32 accessions), subgroup 3 (105 accessions) and subgroup 4 (25 accessions) (Chen et al. 2017). The LD values of the different chromosomes were reported in Chen et al. (2016).

Phenotypic data

The phenotypic values for the wheat starch trait in diverse environments are shown in Table 1. Extensive phenotypic variation for AMS, AMP and TSC among the 205 winter wheat accessions was observed across four environments (i.e., two growing seasons and two locations). The AMS contents ranged from 16.47 to 22.99% in the flour, AMP contents ranged from 38.43 to 61.15%, TSC contents ranged from 55.78 to 82.19%, and AMS/AMP ratio ranged from 33.01 to 52.78%. Broad-sense heritabilities were 89.31, 68.10, 75.36 and 32.45%, respectively, indicating that both genetic and environmental factors influenced the expression of each trait. There was no significant difference found between environments with regard to AMS, AMP and TSC (Table S2).

Thousand kernel weights (TKW) ranged from 26.33 to 60.13 g, and protein contents ranged from 10.30 to 17.98% across environments (Table 2). These two phenotypic data were approximately normally distributed in this population with the absolute values of skewness and kurtosis of less than 1.0. Hence, they belonged to typical quantitative traits controlled by multiple loci.

Marker-trait associations and elite allele exploration

A total of 24,355 mapped SNPs was used for MTA analysis. Forty-seven significant MTAs were detected for all four traits across environments (Table 3, Table S3, Fig. 1). We further analysed MTAs for AMS and AMP by comparing the phenotypic effects

Table 1 Phenotypic values for starch content and starch composition of wheat flour from 205 winter wheat accessions grown in four environments

Trait	Env.	Min. (%)	Max. (%)	Range(%)	Mean (%)	SD	Skewness	Kurtosis	H^2 (%)
AMS	E1	16.56	22.5	5.94	18.45	1.55	0.35	− 1.03	89.31
	E2	16.47	22.99	6.52	19.28	1.59	0.38	− 0.94	
	E3	17.15	22.89	5.74	19.39	1.46	0.49	− 0.93	
	E4	16.47	22.89	6.42	19.93	1.59	0.32	− 1.05	
AMP	E1	39.37	59.51	20.14	51.53	4.1	− 0.44	0.68	68.1
	E2	38.43	59.04	20.61	50.78	4.09	− 0.62	0.97	
	E3	38.43	61.15	22.72	50.95	4.46	− 0.51	0.71	
	E4	39.13	59.51	20.38	49.99	4.14	− 0.49	0.65	
TSC	E1	57.93	81.92	23.99	69.98	5.25	0.05	− 0.28	75.36
	E2	55.78	81.45	25.67	70.06	5.32	− 0.07	− 0.15	
	E3	61.15	82.19	21.04	70.34	5.49	0.04	− 0.24	
	E4	57.71	81.7	23.99	69.92	5.33	0.02	− 0.41	
AMS/AMP	E1	32.97	50.09	17.12	38.02	2.79	1.64	4.35	32.45
	E2	32.63	48.18	15.51	38.07	2.72	1.22	2.56	
	E3	33.43	52.78	19.30	38.22	3.16	2.20	6.26	
	E4	33.01	51.04	18.08	38.03	2.84	1.85	5.76	

E1, E2, E3 and E4, Tai'an 2013; Dezhou 2013; Tai'an 2014; and 2014 Dezhou, respectively; AMS, amylose; AMP, amylopectin; TSC, total starch content

Table 2 Mean phenotypic values for thousand kernel weight (TKW) and grain protein content of 205 winter wheat accessions grown in four environments

Trait	Env.	Min.	Max.	Average ± SD	CV (%)	Skewness	Kurtosis
TKW	E1	28.46	59.63	43.46B ± 5.39	12.37	0.2	0.37
	E2	32.00	60.13	47.89A ± 5.28	11.03	− 0.19	− 0.02
	E3	26.33	56.66	43.45B ± 4.92	11.32	− 0.22	0.29
	E4	29.33	58.03	45.85A ± 4.94	10.77	− 0.24	0.32
Grain protein content	E1	12.13	17.93	14.34A ± 1.02	7.11	0.64	0.74
	E2	12.15	17.95	14.76A ± 1.18	7.99	0.32	− 0.29
	E3	11.30	17.98	14.37A ± 1.01	7.03	0.31	0.63
	E4	10.30	17.05	12.88B ± 1.24	9.63	0.54	0.05

E1, Tai'an 2013; E2, Dezhou 2013; E3, Tai'an 2014; E4, Dezhou 2014

of alleles at each locus to identify elite genes for the starch components and AMS: AMP ratio (Table S4, Table S5). Nine MTAs were recorded for the two starch traits, and there were 11 MTAs for three traits. These SNPs on eight chromosomes, each accounted for 11.26–23.83% of the phenotypic variance. Eighteen MTAs on chromosomes 1B, 2A, 3B, 3D, 4A, 5B, 6A, 6B and 7B were identified as being related to AMS: AMP ratio, each explaining 5.92–17.2% of the phenotypic variation. Nine MTAs were detected in two environments; seven in E1 and E2, and two in E3 and E4.

Fifteen MTAs for AMS were identified on chromosomes 2A, 2B, 3A and 4A explaining 11.8–18.41% of the phenotypic variation. Two MTAs, *IAAV4464* (2A_112) and *JD_c3742_1130* (2A_112), on chromosome 2A were detected in three environments; these MTAs located at the same position had the highest R^2 (18.41%) and smallest P values (Fig. 1b). Twelve of the 15 MTAs showed significant phenotypic differences among alleles ($P < 0.01$; Table S4), and the same MTAs exhibited phenotypic differences in environments E2 and E4 ($P < 0.05$). Alleles A and G of marker *Kukri_c5615_1214* (3A_93) were

Table 3 Main marker-trait associations detected in at least two of four environments

Trait	Marker	Chr.	Position (cM)	*P* value				R^2 (%)			
				E1	E2	E3	E4	P1	P2	P3	P4
AMS	IAAV4464	2A	112	6.70E−04	2.84E−05		5.34E−05	12.04	18.41		17.12
	JD_c3742_1130	2A	112	6.70E−04	2.84E−05		5.34E−05	12.04	18.41		17.12
	Kukri_c50842_573	2B	104		5.42E−04		2.85E−04		12.39		13.69
	TA005830-0667	2B	111		1.70E−04	1.10E−04			14.71	15.69	
	RAC875_c6280_292	4A	127		1.27E−04	1.82E−04			15.32	14.59	
AMP	IAAV4464	2A	112	5.14E−04	9.56E−05		5.60E−04	12.58	15.9		12.33
	JD_c3742_1130	2A	112	5.14E−04	9.56E−05		5.60E−04	12.58	15.9		12.33
	Kukri_c50842_573	2B	104	8.88E−04	1.00E−05	1.45E−04	6.68E−06	11.48	20.61	15.12	21.51
	TA004152-0921	2B	107	9.00E−04	1.20E−05	1.05E−04	5.91E−06	11.45	20.23	15.78	21.77
	Kukri_c5615_1214	3A	93		3.19E−04		1.75E−04		13.45		14.69
	RAC875_c6280_292	4A	127	1.92E−04	1.27E−04		1.07E−04	14.59	15.32		15.69
	JD_c6831_221	6A	141	5.90E−04			9.64E−04	12.3			11.26
	Excalibur_c16376_351	6B	0	1.18E−04	1.33E−04	4.96E−06	2.65E−05	15.68	15.29	22.44	18.71
TSC	IAAV4464	2A	112	2.10E−04	2.20E−05		1.20E−04	14.37	18.96		15.51
	JD_c3742_1130	2A	112	2.10E−04	2.20E−05		1.20E−04	14.37	18.96		15.51
	Kukri_c50842_573	2B	104		9.39E−06	1.60E−04	4.50E−06		20.75	14.92	22.35
	TA004152-0921	2B	107		1.36E−05	1.10E−04	4.90E−06		19.96	15.7	22.18
	Kukri_c5615_1214	3A	93		1.30E−04		7.20E−05		15.31		16.49
	RAC875_c6280_292	4A	127	1.70E−04	1.10E−04	5.20E−04	1.20E−04	14.9	15.58	12.51	15.39
	JD_c6831_221	6A	141	5.50E−04	7.20E−04		5.60E−04	12.46	11.82		12.34
	Excalibur_c16376_351	6B	0	1.10E−04	9.30E−05	2.65E−06	3.10E−05	15.91	16.03	23.83	18.35
AMS/AMP	BS00021739_51	2A	106	3.57E−05	3.57E−05			11.45	11.45		
	BS00110445_51	3B	71	2.03E−05	2.03E−05			12.05	12.05		
	Tdurum_contig10408_1548	3D	143	1.64E−05	1.64E−05			12.33	12.33		
	IAAV1943	4A	144	2.98E−05	2.98E−05			11.67	11.67		
	BS00090253_51	6A	13	1.15E−05	1.15E−05			13.16	13.16		
	BS00027770_51	6B	98	1.14E−05	1.14E−05			12.82	12.82		
	Tdurum_contig48824_476	7B	54			2.86E−04	4.77E−05			7.01	9.17

E1, E2, E3 and E4, Tai'an 2013; Dezhou 2013; Tai'an 2014; and 2014 Dezhou, respectively; AMS, amylose; AMP, amylopectin; TSC, total starch content

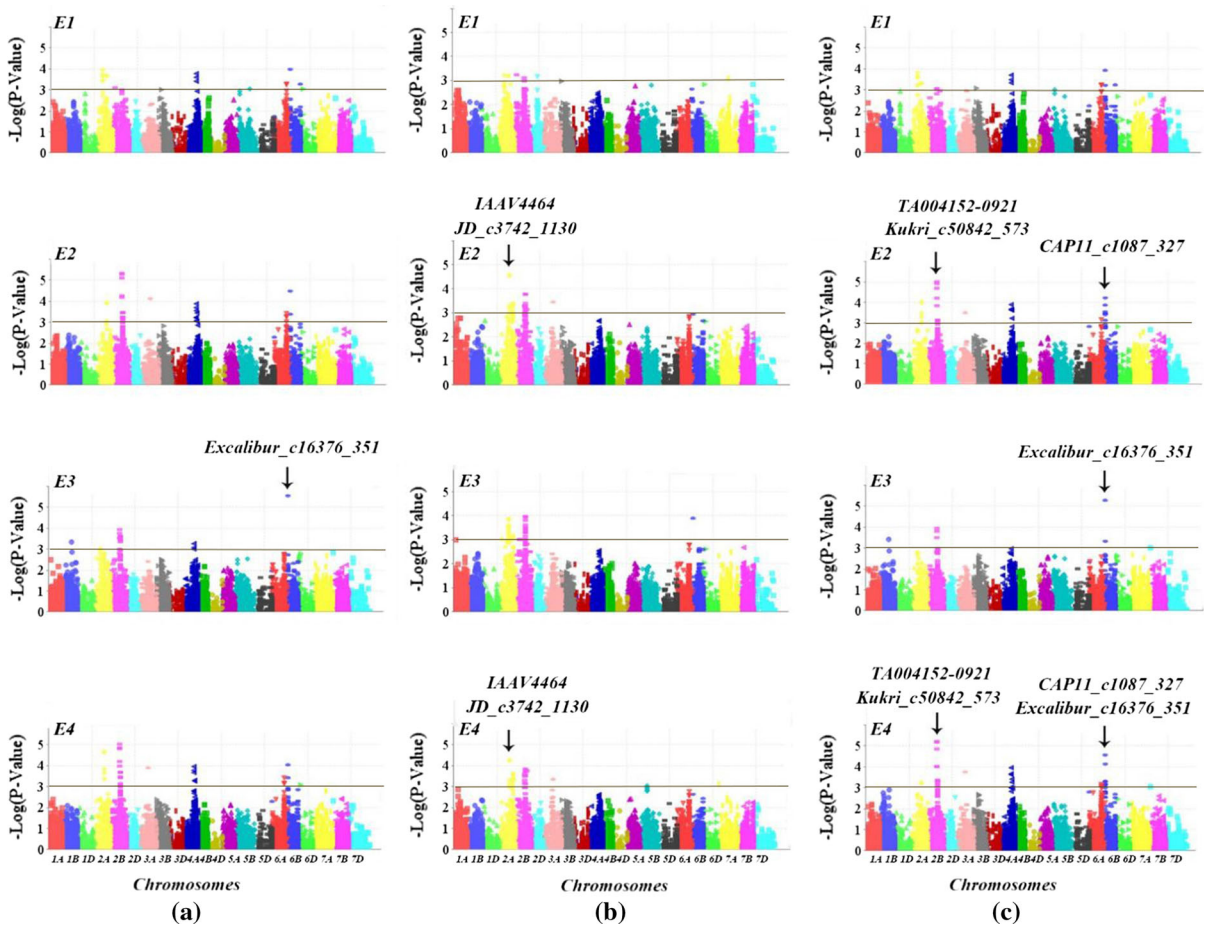

Fig. 1 Manhattan plots of GWAS for three traits with a mixed linear model. a–c indicate Manhattan plots for total starch content, amylose and amylopectin, respectively; E1, E2, E3 and E4, Tai'an 2013; Dezhou 2013; Tai'an 2014; and 2014 Dezhou, respectively

associated with the largest phenotypic differences (5.80%). The phenotypic value of AMS associated with *Kukri_c5615_1214-A* (3A_93) was significantly higher than that associated with *Kukri_c5615_1214-G* (3A_93) across all four environments, indicating that *Kukri_c5615_1214-A* (3A_93) was a more elite allele than *Kukri_c5615_1214-G* (3A_93). The marker *RFL_Contig4517_1276* (2A_110) revealed no significant phenotypic differences among its alleles.

Twenty-three MTAs for AMP detected on chromosomes 2A, 2B, 3A, 3B, 4A, 6A, 6B and 7D accounted for 11.26 to 22.44% of the phenotypic variation. Among them, six MTAs were detected in three environments. Markers *Kukri_c50842_573* (2B_104) and *TA004152-0921* (2B_107) on chromosome 2B and *Excalibur_c16376_351* (6B_0) and *CAP11_c1087_327* (6B_6) on chromosome 6B were identified in all four environments and, except for *Excalibur_c16376_351* (6B_0) in E3 and E4, revealed significant MTAs in E2 and E4 (Fig. 1c). Twenty-two MTAs for AMP showed significant allelic differences in phenotype ($P < 0.01$). Among them, *Kukri_c5615_1214* (3A_93) had the most significant effect, increasing AMP by 15.32%, and allele G was identified as an elite allele. Other elite alleles at each locus increased AMP from 1.25 to 4.43%. Eight MTAs for both AMS and AMP were identified; of these, six exhibited highly significant phenotypic differences between alleles for both traits.

Twenty-two MTAs for TSC were detected on chromosomes 2A, 2B, 3A, 3B, 4A, 6A and 6B, explaining 11.31 to 23.83% of the phenotypic variation. Four MTAs were found in four environments and 12 MTAs were detected in three environments.

Table 4 Significant ($P \leq 10^{-4}$) or stable MTAs and percentages of phenotypic variation explained for mean thousand kernel weights from four environments

Trait	Marker	Chr.	Position	R^2 (%)	*P* value
TKW	CAP12_c4704_232	4B	85	5.74(E2),6.81(E3)	6.34E−04,8.56E−04
	BS00023893_51	6A	86	7.32(E1),9.86(E2),8.13(E4)	3.47E−04,5.03E−04,6.27E−04
	wsnp_BQ161779D_Ta_2_1	6D	86	7.79(E2)	9.91E−05
	RAC875_c17479_359	3A	93	11.28(E2)	9.37E−05
Grain protein content	Tdurum_contig4974_355	4B	61	9.67(E2),7.10(E4)	1.56E−05,3.62E−04
	Excalibur_c47675_176	6B	91	8.11(E4)	0.0000674
	wsnp_Ex_c14654_22713386	7A	42	10.88(E2),8.14(E4)	5.43E−06,6.99E−05
	D_contig06359_118	7D	56	10.62(E2),7.40(E4)	4.24E−06,2.19–04
	Excalibur_c1208_72	5A	62	6.75(E1),6.93(E2)	2.64E−04,2.03E−04

Marker *Excalibur_c16376_351* (6B_0) had the highest R^2 (23.83%) and was significant in four environments (Fig. 1a).

Eleven of 18 MTAs for AMS: AMP ratio (Table S5) exhibited significant phenotypic differences among alleles ($P < 0.05$) in at least two environments; six showed significant phenotypic differences between alleles ($P < 0.01$), and three showed highly significant differences in two environments. Markers *BS00022255_51* (1B_57) and *D_contig25392_201* (1B_61) had the most significant effects, increasing AMP by 3.77%. The T allele of the former marker and A allele of the latter were elite alleles.

We identified four significant MTAs for TKW on chromosomes 3A, 4B, 6A and 6D, explaining 5.74 to 11.28% of the phenotypic variation, and SNP locus *BS00023893_51* (6A_86) on chromosome 6A was identified in three environments (Table 4). For grain protein content, five MTAs were found on chromosomes 4B, 6B, 7A, 7D and 5A, explaining 6.75 to 10.88% of the phenotypic variation (Table 4).

Putative candidate genes linked to starch-related traits

Significant MTAs identified in more than two environments and correlated with more than one trait were selected for candidate gene prediction (Table 5). For marker *IAAV4464* on chromosome 2AL there were four candidates but gene *TRIAE_CS42_2AL_TGACv1_093900_AA0288950* was related to beta-glucosidase and hydrolysis of O-glycosyl compounds that participate in carbohydrate metabolism. The candidate gene for marker *JD_c3742_1130* on 2AL also participates in carbohydrate metabolism. Significant markers *RAC875_c6280_292* and *Tdurum_contig41127_265* on chromosome 4AL had the same candidate in gene *TRIAE_CS42_4AL_TGACv1_288945_AA0961860* that is expressed in the aleurone layer and endosperm 10–30 days post anthesis. The candidate gene *TRIAE_CS42_2AL_TGACv1_093900_AA0288950* for marker *BobWhite_c10583_352* was predicted to participate in carbohydrate metabolism. Both genes are novel with unknown function, and are different from *Wx-B1*. Markers *Excalibur_c16376_351* and *CAP11_c1087_327* had the same candidate gene related to adenosine diphosphate (ADP) binding. These candidate wheat genes could be related to starch synthesis; their functions will be investigated in future research.

Discussion

A number of significant loci were identified in this study, suggesting the presence of at least some MTAs with medium and small effects on starch traits. Presumably, we should focus on highly significant or stable MTAs and multi-trait MTAs. Due to differences in marker types and marker positions on different genetic maps, the MTA results of our study were extensively compared with previously reported MTA results involving the same chromosome arms.

In previous QTL mapping studies, at least 10 chromosomes related to the TSC were identified, namely, 1A, 1B, 1D, 2A, 3D, 3B, 4A, 5B, 5D and 7D.

Table 5 Predicted candidate genes for SNP markers significantly associated with amylose (AMS), amylopectin (AMP) and total starch content (TSC) in more than two environments

Marker	Chromosome	Candidate gene	Function	Biological process/Expression	Species	
IAAV4464	2AL	GeneID:4336389	Beta-glucosidase 16 isoform X1		*Oryza sativa ssp. japonica*	
		BGLU45	Beta-glucosidase 45		*Arabidopsis thaliana*	
		GeneID:109763016	GDT1-like protein 5 isoform X2		*Aegilops tauschii ssp. tauschii*	
		TRIAE_CS42_2AL_TGACv1_093900_AA0288950	Hydrolysing O-glycosyl compounds	Carbohydrate metabolism	*Triticum aestivum*	
JD_c3742_1130	2AL	GeneID:9270188	F-box/kelch-repeat protein		*Oryza sativa ssp. japonica*	
		TRIAE_CS42_2AL_TGACv1_093900_AA0288950	Hydrolysing O-glycosyl compounds	Carbohydrate metabolism	*Triticum aestivum*	
wsnp_Ex_c63909_62932437	2AL	TRIAE_CS42_2AL_TGACv1_094920_AA0304860	N-acetyltransferase activity; metal ion binding	10–30 days post anthesis; expressed in aleurone layer and endosperm	*Triticum aestivum*	
RFL_Contig4517_1276	2AL	TRIAE_CS42_2AL_TGACv1_094920_AA0304860	N-acetyltransferase activity; metal ion binding	10–30 days post anthesis; expressed in aleurone layer and endosperm	*Triticum aestivum*	
RAC875_c6280_292	4AL	TRIAE_CS42_4AL_TGACv1_288945_AA0961860		10–30 days post anthesis; expressed in aleurone layer and endosperm	*Triticum aestivum*	Novel gene
Tdurum_contig41127_265	4AL	TRIAE_CS42_4AL_TGACv1_288945_AA0961860		10–30 days post anthesis; expressed in aleurone layer and endosperm	*Triticum aestivum*	Novel gene
BobWhite_c10583_352	4AL	TRIAE_CS42_2AL_TGACv1_093900_AA0288950		Carbohydrate metabolism	*Triticum aestivum*	Novel gene
Excalibur_c16376_351	6BS	GeneID:4350815	RPP13-like protein 3		*Oryza sativa ssp. Japonica*	
		TRIAE_CS42_6BS_TGACv1_513154_AA1632250		ADP binding	*Triticum aestivum*	Novel gene
CAP11_c1087_327	6BS	TRIAE_CS42_6BS_TGACv1_513154_AA1632250		ADP binding	*Triticum aestivum*	Novel gene

Eight chromosomes had QTL for AMS content, including 1B, 2A, 2D, 3A, 3B, 4A, 5D and 7D, and eight chromosomes were identified for AMP content, namely 1B, 2A, 2B, 3A, 3B, 4D, 5A and 5D (McCartney et al. 2006; Sun et al. 2008; Tian et al. 2015; Deng et al. 2015). Starch granule size was related to AMS and AMP (Peterson and Fulcher 2001), that is, a higher amylose content in larger granules was found than that in smaller granules, while the amylopectin content of small starch granules was higher than that of large starch granules, so A-granules had a higher ratio of amylose to amylopectin (Park et al. 2004; Li et al. 2011). Li et al. (2017b) identified 15 chromosomes, including 2A, 2B, 3A, 4A, 6A, 6B and 7D, that were related to the percentage volumes of A- and B-granules and the ratio of A-/B-granule volumes. In the present study, seven chromosomes were linked to TSC (2A, 2B, 3A, 3B, 4A, 6A and 6B), chromosomes 2A, 2B and 3A were linked to AMS, and eight chromosomes were linked to AMP (2A, 2B, 3A, 3B, 4A, 6A, 6B and 7D). Thus, chromosomes 2A, 2B and 3A were associated all three traits (AMS, AMP and TSC), and were also related to A- and B-granules and starch components in previous studies. No associations on chromosomes 1A, 1B, 1D, 3D, 5A, 5B and 5D were identified in this study, but chromosomes 4A and 7D were linked to AMP but not AMS, and chromosomes 6A and 6B were related to AMP and TSC. It is interesting that chromosomes 3A and 6D were also related to TKW, and chromosomes 6B and 7D were related to grain protein content. These results indicated that starch content was related to TKW and grain protein content. Starch development also affects grain yield (Hurkman et al. 2003; Tetlow et al. 2004). Moreover, there was a negative correlation between starch content and grain protein content. It is possible that some genes on the same chromosome affect starch content, TKW and grain protein content.

Using association analysis to identify elite alleles has become a useful strategy for plant genomics research (Cai et al. 2014; Li et al. 2012). In this study, we established a link between genotypes and AMS and AMP phenotypes by analysing differences in phenotypic values among various alleles and identified elite alleles for AMS and AMP. For example, the allele *Kukri_c5615_1214-G* (3A_93) increased AMS and AMP by 5.80 and 15.32%, respectively, and the allele *Excalibur_c16376_351-T* (6B_0) increased AMP by 4.43%. Consequently, lines that carry these elite alleles could be used as parents for breeding. The marker *Kukri_c5615_1214* (3A_93) on chromosome 3A was a stable multi-effect MTA. Its elite allele had a considerable effect in increasing both AMS and AMP contents. These results indicated that these markers were closely related to genes involved in starch synthesis.

GBSS is a key enzyme in AMS synthesis. The genes expressing GBSS are located on 7AS, 4AL and 7DS (Murai et al. 1999; Yan et al. 2007). Previous studies found QTLs for AMS content near the *Wx-B1* on chromosome 4A (Araki et al. 1999). During grain development, *QTsc-4A.1* and *QAms-4A.1* located in the *Xwmc262–Xbarc343* interval made a large contribution to TSC and AMS synthesis over the whole grain-filling process (Tian et al. 2015). And by comparing them with physical map (Cui et al., 2014), *QTsc-4A.1* and *QAms-4A.1* were found on chromosome 4AL. In the present study, three SNP markers (*RAC875_c6280_292*, *Tdurum_contig41127_265* and *BobWhite_c10583_352*) were also detected on chromosome 4A and were closely associated with all three traits. These three SNP markers were associated with unknown genes functioning in carbohydrate metabolism in the aleurone layer and endosperm at 10–30 days post-anthesis. By prediction and comparison these genes seemed to be new (Table 5) because of their different positions from the *Wx-B1* locus.

Compared to previous studies, chromosomes 2A, 3A and 6B appeared to be important in control of starch synthesis in the present study. The SNPs *IAAV4464 (2A_112)* and *JD_c3742_1130 (2A_112)* on chromosome 2A were related to all three traits (TSC, AMS and AMP). Annotations of the candidate genes predicted participation in carbohydrate metabolism (Table 5). The marker *Excalibur_c16376_351* (6B_0) on chromosome 6B had the highest R^2 (23.83% for TSC and 22.44% for AMP) and was detected in all four environments. The associated gene was predicted to be a novel gene for starch synthesis (Table 5). Chromosome 6B appears to be an essential chromosome for control of components of starch content.

Conclusions

Thirty-two significant marker-trait associations (MTAs) were detected for total starch (TSC), amylose

(AMS) and amylopectin (AMP) contents in four environments. Fourteen MTAs were detected for two traits, and eight MTAs were found for all three traits. The SNPs were distributed across seven chromosomes, 2A, 2B, 3A, 3B, 4A, 6A and 6B. A set of elite alleles was identified, including *Kukri_c5615_1214-A*, *Excalibur_c16376_351-T*, *BS00022255_51-T* and *D_contig25392_201-T*. Fourteen candidate genes associated with significant markers were identified. The lines that carry these elite alleles could be used as parents in wheat breeding. These results could lay the foundation for fine mapping, gene discovery, and molecular marker-assisted selection of these three traits in wheat.

Acknowledgements We are grateful for grants from the Natural Science Foundation of China (No. 31871613), the key research and development plan of Shandong Province (2017GNC10102), the University of Science and Technology of Shandong Province (J17KA148), and the Shandong "Double Tops" Program.

Author contribution ZD designed and revised this paper; XC wrote the manuscript; WF analysed the data; XC, SX, YJ and SS investigated the phenotypic data; GC constructed the map; and JT reviewed the manuscript. All authors have read and approved the paper.

Compliance with ethical standards

Conflicts of Interest The authors declare that they have no conflicts of interest.

References

Araki E, Miura H, Sawada S (1999) Identification of genetic loci affecting amylose content and agronomic traits on chromosome 4A of wheat. Theor Appl Genet 98:977–984

Bandillo N, Jarquin D, Song QJ, Nelson R, Cregan P, Specht J, Specht J, Lorenz A (2015) A population structure and genome-wide association analysis on the USDA soybean germplasm collection. Plant Genome 8:832

Batey IL, Hayden MJ, Cai S, Sharp PJ, Cornish GB, Morell MK, Appels R (2002) Genetic mapping of commercially significant starch characteristics in wheat crosses. Aust J Agric Res 52:1287–1296

Bradbury PJ, Parker T, Hamblin MT, Jannink JL (2011) Assessment of power and false discovery rate in genome-wide association studies using the barley CAP germplasm. Crop Sci 51:52–59

Caballero L, Bancel E, Debiton C, Branlard G (2008) Granule-bound starch synthase (GBSS) diversity of ancient wheat and related species. Plant Breed 127:548–553

Cai CP, Ye WX, Zhang TZ, Guo WZ (2014) Association analysis of fiber quality traits and exploration of elite alleles in upland cotton cultivars/accessions (*Gossypium hirsutum* L.). J Integr Plant Biol 56:51–62

Chen GF, Chen JS, Tian JC (2015) Genome-wide association analysis between SNP markers and plant height related traits in wheat. Acta Agron Sin 41:1500

Chen GF, Zhang H, Deng ZY, Wu RG, Li DM, Wang MY, Tian JC (2016) Genome-wide association study for kernel weight-related traits using SNPs in a Chinese winter wheat population. Euphytica 212:173–185

Chen GF, Wu RG, Li DM, Yu HX, Deng ZY, Tian JC (2017) Genome-wide association study for seedling emergence and tiller number using SNP markers in an elite winter wheat population. J Genet 96:177–186

Cho YH, Shim JY, Lee HG (2007) Characteristics of wheat flour dough and noodles with amylopectin content and hydrocolloids. Korean J Food Sci Technol 39:138–145

Cockram J, White J, Zuluaga DL, Smith D, Comadran J, Macaulay M, Luo ZW, Kearsey MJ, Werner P, Harrap D, Tapsell C, Liu H, Hedley PE, Stein N, Schulte D, Steuernagel B, Marshall DF, Thomas WTB, Ramsay L, Mackay I, Balding DJ, Waugh R, O'Sullivan DM, Booer C, Pike S, Hamilton G, Jellis G, Davies N, Ross A, Bury P, Habgood R, Klose S, Vequaud D, Christerson T, Brosnan J, Newton A, Russell J, Shaw P, Bayles R, Wang MH (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. Proc Natl Acad Sci U S A 107:21611–21616

Crofts N, Nakamura Y, Fujita N (2017) Critical and speculative review of the roles of multi-protein complexes in starch biosynthesis in cereals. Plant Sci 262:1–8

Cui F, Fan XL, Zhao CH, Zhang W, Chen M, Ji J, Li JM (2014) A novel genetic map of wheat: utility for mapping QTL for yield under different nitrogen treatments. BMC Genet 15:57

Deng ZY, Hu SN, Chen F, Li WJ, Chen JS, Sun CL, Zhang YX, Wang SY, Song XJ, Tian JC (2015) Genetic dissection of interaction between wheat protein and starch using three mapping populations. Mol Breed 35:12

Deng ZY, Cui Y, Han QD, Fang WQ, Li JF, Tian JC (2017) Discovery of consistent QTLs of wheat spike-related traits under nitrogen treatment at different development stages. Front Plant Sci 8:2120

Deng ZY, Fang WQ, Guo X, Zhao XY, Guo H, Hu SN, Tian JC (2018) Genetic dissection of interactions between wheat flour starch and its components in two populations using two QTL mapping methods. Mol Breed 38:41

Gupta PK, Mir RR, Mohan A, Kumar J (2008) Wheat genomics: present status and future prospects. Int J Plant Genom 2008:896451

Huang XH, Zhao Y, Wei XH, Li CY, Wang AH, Zhao Q, Li WJ, Guo YL, Deng LW, Zhu CR, Fan DL, Lu YQ, Weng QJ, Liu KY, Zhou TY, Jing YF, Si LZ, Dong GJ, Huang T, Lu TT, Feng Q, Qian Q, Li JY, Han B (2011) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet 44:32–39

Hurkman WJ, McCue KF, Altenbach S, Korn A, Tanaka CK, Kothari KM, Johnson EL, Bechtel DB, Wilson JD, Anderson OD, DuPont FM (2003) Effect of temperature on

expression of genes encoding enzymes for starch biosynthesis in developing wheat endosperm. Plant Sci 164:873–881

Jin YH, Zhang KL, Zhang XC, Du JH (2009) Determination of amylose and amylopectin in wheat and wheat malt by dual -wavelength spectrophotometry. J Chin Cereals Oils Assoc 24:137–140

Li ZY, Chu XS, Mouille G, Yan LL, Kosar-Hashemi B, Hey S, Napier J, Shewry P, Clarke B, Appels R, Morell MK, Rahman S (1999) The localization and expression of the class ii starch synthases of wheat. Plant Physiol 120:1147–1155

Li ZY, Sun F, Xu SM, Chu XS, Mukai Y, Yamamoto M, Ali S, Rampling L, Kosar-Hashemi B, Rahman S, Morell M (2003) The structural organization of the gene encoding class II starch synthase of wheat and barley and the evolution of the genes encoding starch synthases in plants. Funct Integr Genom 3:76–85

Li WY, Yin YZ, Shi XQ, Yan SH, Wang ZL (2011) Comparison of starch composition and pasting properties between A and B type starch granule in wheat grain. Acta Agric Boreali Sin 26:136–139

Li WY, Zhang B, Zhang JN, Cheng XP, Li RZ, Jing RL (2012) Exploring elite alleles for chlorophyll content of flag leaf in natural population of wheat by association analysis. Theor Appl Genet 38:962–970

Li GQ, Bai GH, Carver BF, Elliott NC, Bennett RS, Wu YQ, Hunger R, Bonman JM, Xu XY (2017a) Genome-wide association study reveals genetic architecture of coleoptile length in wheat. Theor Appl Genet 130:391–401

Li JY, Rasheed A, Guo Q, Dong Y, Liu JD, Xia XC, Zhang Y, He ZH (2017b) Genome-wide association mapping of starch granule size distribution in common wheat. J Cereal Sci 77:211–218

Liu JJ, He ZH, Zhao ZD, Pena RJ, Rajaram S (2003) Wheat quality and quality parameters of cooked dry white Chinese noodles. Euphytica 131:147–154

Liu N, Xue YD, Guo ZY, Li WH, Tang JH (2016) Genome-wide association study identifies candidate genes for starch content regulation in maize kernels. Front Plant Sci 7:1046

Martin C, Smith AM (1995) Starch biosynthesis. Plant Cell 7:971–985

Martin JM, Talbert LE, Habernicht DK, Lanning SP, Sherman JD, Carlson G, Giroux MJ (2004) Reduced amylose effects on bread and white salted noodle quality. Cereal Chem 81:188–193

Mccartney CA, Somers DJ, Lukow O, Ames N, Noll J, Cloutier S (2006) QTL analysis of quality traits in the spring wheat cross RL4452 × 'AC Domain'. Plant Breed 125:565–575

Mick E, McGough JJ, Middleton FA, Neale B, Faraone VS (2011) Genome-wide association study of blood pressure response to methylphenidate treatment of attention-deficit/hyperactivity disorder. Prog Neuro-Psychopharmacol Biol Psychiatry 35:466–472

Murai J, Taira T, Ohta D (1999) Isolation and characterization of the three waxy genes granule-bound starch synthase in hexaploid wheat. Gene 234:71–79

Nakamura T, Yamamori M, Hirano H, Hidaka S (1993) Identification of three Wx proteins in wheat (*Triticum aestivum* L.). Biochem Genet 31:75–86

Nakamura Y, Ono M, Sawada T, Crofts N, Fujita N, Steup M (2017) Characterization of the functional interactions of plastidial starch phosphorylase and starch branching enzymes from rice endosperm during reserve starch biosynthesis. Plant Sci 2017:83–95

Newell MA, Asoro FG, Scott MP, White PJ, Beavis WD, Jannink JL (2012) Genome-wide association study for oat (*Avena sativa* L.) beta-glucan concentration using germplasm of worldwide origin. Theor Appl Genet 125:1687–1696

Ovenden B, Milgate A, Wade LJ, Rebetzke GJ, Holland JB (2017) Genome-wide associations for water-soluble carbohydrate concentration and relative maturity in wheat using SNP and DArT marker arrays. G3 (Bethesda, Md) 7:2821–2830

Park SH, Wilson JD, Chung OK, Seib PA (2004) Size distribution and properties of wheat starch granules in relation to crumb grain score of pup-loaf bread. Cereal Chem 81:699–704

Peterson DG, Fulcher RG (2001) Variation in Minnesota HRS wheat: starch granule size distribution. Food Res Int 34:357–363

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516–1517

Sajjad M, Khan SH, Kazi AM (2012) The low down on association mapping in hexaploid wheat (*Triticum aestivum* L.). J Crop Sci Biotechnol 15:147–158

Shu XL, Backes G, Rasmussen SK (2012) Genome-wide association study of resistant starch (RS) phenotypes in a barley variety collection. J Agric Food Chem 60:10302–10311

Song JM, Liu AF, Li HS, Dai S, Liu JJ, Zhao ZD, Liu GT (2008) Relationship between starch physiochemical properties of wheat grain and noodle quality. Sci Agric Sin 41:272–279

Sukumaran S, Yu J (2014) Association mapping of genetic resources: achievements and future perspectives. In: Roberto T, Andreas G, Emile F (eds) Genomics of plant genetic resources. Springer, Netherlands, pp 207–235

Sun HY, Lv JH, Fan YD, Zhao Y, Kong FM, Wang HG, Li SS (2008) QTL analysis of wheat protein and starch quality. Prog Nat Sci 5:505–513

Tetlow IJ, Morell MK, Emes MJ (2004) Recent developments in understanding the regulation of starch metabolism in higher plants. J Exp Bot 55:2131–2145

Tian B, Deng ZY, Xie QF, Tian JC (2015) Genetic dissection of the developmental behaviour of total starch content and its components in wheat grain. Crop Pasture Sci 66:445–455

Wang XZ, Li H, Gou HJ, Zhang L (1998) The relativity between processing characteristics of wheat and quality characteristics of steamed bread. J Chin Cereals Oils Assoc 6:6–8

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, 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

Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES IV (2004) Dissection of maize kernel composition and starch production by candidate gene associations. Plant Cell 16:2719–2733

Yan SH, Wang ZL, Dai ZM, Li WY, Fu GZ, He MR, Yin YP (2007) Comparison of starch synthase activity and starch accumulation in two different wheat varieties with different amylose contents. Acta Agron Sin 33:84–89

Yu JM, Buckler ES (2006) Genetic association mapping and genome organization of maize. Curr Opin Biotechnol 17:155–160

Zhou ZM (2012) Effect of wheat starch on the quality of steamed bread. Farm Prod Process 3:7

Zhu CS, Gore M, Buckler ES, Yu JM (2008) Status and prospects of association mapping in plants. Plant Genome 1:5–20