Mol Breeding (2023) 43:91
https://doi.org/10.1007/s11032-023-01437-6

Identification of novel loci associated with starch content in maize kernels by a genome-wide association study using an enlarged SNP panel

Haiyang Duan · Jianxin Li · Li Sun · Xuehang Xiong · Shuhao Xu · Yan Sun · Xiaolong Ju · Zhengjie Xue · Jionghao Gao · Yan Wang · Huiling Xie · Dong Ding · Xuehai Zhang · Jihua Tang

Received: 5 June 2023 / Accepted: 24 November 2023 / Published online: 12 December 2023

Abstract Starch is a major component of cereals, comprising over 70% of dry weight. It serves as a primary carbon source for humans and animals. In addition, starch is an indispensable industrial raw material. While maize (*Zea mays*) is a key crop and the primary source of starch, the genetic basis for starch content in maize kernels remains poorly understood. In this study, using an enlarged panel, we conducted a genome-wide association study (GWAS) based on best linear unbiased prediction (BLUP) value for starch content of 261 inbred lines across three environments. Compared with previous study, we identified 14 additional significant quantitative trait loci (QTL), encompassed a total of 42 genes, and indicated that increased marker density contributes to improved statistical power. By integrating gene expression profiling, Gene Ontology (GO) enrichment and haplotype analysis, several potential target

Haiyang Duan and Jianxin Li contributed equally to this work.

Highlights
1. By utilizing an enlarged SNP panel, we identified 14 novel loci associated with starch content, highlighting the importance of increased marker density in improving statistical power.
2. The candidate gene *ZmAPC4* encodes a protein belonging to the WD40 repeat-like superfamily and exhibits high expression in maize endosperm, it plays a pivotal role as a regulator in the synthesis of starch in maize kernel.
3. As a notable achievement, we have successfully developed molecular markers that can effectively distinguish maize inbred lines based on their starch content.
4. Our findings provide a valuable reference for enhancing starch content to generate more bioenergy and have the potential to contribute to the advancement of more productive and sustainable agricultural practices.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11032-023-01437-6.

H. Duan · J. Li · L. Sun · X. Xiong · S. Xu · Y. Sun · X. Ju · Z. Xue · J. Gao · H. Xie · D. Ding · X. Zhang · J. Tang
National Key Laboratory of Wheat and Maize Crops Science, College of Agronomy, Henan Agricultural University, Zhengzhou, China
e-mail: 18839046763@163.com

J. Li
e-mail: bnmvbncvb@163.com

L. Sun
e-mail: sun_lier@126.com

X. Xiong
e-mail: xuehang_xiong@163.com

S. Xu
e-mail: 508591485@qq.com

Y. Sun
e-mail: sunyan15938845165@126.com

X. Ju
e-mail: juxiaolog@163.com

Z. Xue
e-mail: 2653271661@qq.com

genes that may play a role in regulating starch content in maize kernels have been identified. Notably, we found that *ZmAPC4*, associated with the significant SNP chr4.S_175584318, which encodes a WD40 repeat-like superfamily protein and is highly expressed in maize endosperm, might be a crucial regulator of maize kernel starch synthesis. Out of the 261 inbred lines analyzed, they were categorized into four haplotypes. Remarkably, it was observed that the inbred lines harboring hap4 demonstrated the highest starch content compared to the other haplotypes. Additionally, as a significant achievement, we have developed molecular markers that effectively differentiate maize inbred lines based on their starch content. Overall, our study provides valuable insights into the genetic basis of starch content and the molecular markers can be useful in breeding programs aimed at developing maize varieties with high starch content, thereby improving breeding efficiency.

Keywords Starch content · Enlarged SNP panel · Genome-wide association study · *ZmAPC4* · Molecular markers

Introduction

Starch is a crucial energy source that has played a significant role in human social activities (Kumar et al. 2021). Crop grains, which contain large amounts of starch, are utilized for food, animal feed, biofuels, and other products (Wu et al. 2021). With global maize (*Zea mays*) yield surpassing 1.086 billion tons in 2022, starch has become more accessible than ever before. However, rapid population growth and a deteriorating ecological environment have placed enormous pressure on food security (Dossa et al. 2021). Currently, increasing yield per unit area has reached a bottleneck, and providing additional arable land to meet the demands of a growing population is not feasible (Yin et al. 2020). Therefore, it's necessary to increase the accumulation of crop dry matter to generate more energy. The endosperm accounts for over 90% of the total grain weight, and the accumulation of starch typically begins in the central region of the endosperm, progressing to the aleurone layer, to form the starchy endosperm in mature grains (Wu et al. 2016). Ensuring the development of endosperm and promoting more starch accumulation is thus essential.

Starch plays a crucial role in plant development. The starch stored in the endosperm provides nutrition for embryo development and serves as an energy source during seed germination (Liu et al. 2022). Starch biosynthesis involves a series of enzymatic reactions, with five classes of enzymes playing a particularly important role: adenosine-diphosphate-glucose pyrophosphorylases (AGPases), soluble starch synthases (SSs), granule-bound starch synthases (GBSSs), branching enzymes (BEs), and debranching enzymes (DBEs). Among these enzymes, AGPases are rate-limiting and can slow down starch synthesis, while GBSSs and SSs are responsible for the biosynthesis of amylose and amylopectin, respectively (Tetlow 2011). BEs and DBEs, on the other hand, assist in breaking down abnormal glucans. Additionally, endosperm development is regulated by several key cell cycle regulators, including *RBR* protein, CDK/cyclin complex, CDK-specific inhibitor, and APC/C (Cross and Umen 2015). Downregulating *RBR1* promotes mitosis and increases cell number, while overexpressing type A CDK leads to reduced accumulation of storage material by preventing endonuclear replication (Leiva-Neto et al. 2004; Sabelli et al. 2013). *KRP* is another important cell cycle regulator involved in endosperm development, it binds A-type CDKs and D-type cyclins to form a complex (Dante et al. 2014).

Several transcription factors that regulate starch content in maize endosperm have been documented. Opaque2 (*O2*) and prolamine-box binding factor

J. Gao
e-mail: gao13598069385@126.com

H. Xie
e-mail: xiehuilingtang@sohu.com

D. Ding
e-mail: dingdong0216@hotmail.com

Y. Wang
Zhucheng Mingjue Tender Company Limited, Weifang, China
e-mail: zcmjzb@126.com

X. Zhang (✉)
Department of Agronomy, Henan Agricultural University, Agricultural Road No. 63, Zhengzhou 450002, China
e-mail: xuehai85@126.com

J. Tang (✉)
The Shennong Laboratory, Zhengzhou, China
e-mail: tangjihua1@163.com

(*PBF*) are noteworthy examples as they not only govern the synthesis of storage protein zein but also exert control over starch synthesis (Zhang et al. 2016a). Another transcription factor, *ZmbZIP22*, exhibits specific expressed in maize endosperm, and it is known to bind to the promoter of the 27-kD γ-zein gene and plays a pivotal role in regulating its expression. Notably, the overexpression of *ZmbZIP22* has been shown to reduce the size of starch granules, indicating its role as a negative regulator of starch synthesis (Dong et al. 2019; Li et al. 2018a). Additionally, two endosperm-specific NAC transcription factors, *ZmNAC128* and *ZmNAC130*, have been identified. Downregulation of the expression of *ZmNAC128* and *ZmNAC130* results in smaller kernels, and a reduction in starch content. Intriguingly, these two factors, along with *O2*, synergistically promote endosperm filling (Chen et al. 2023; Zhang et al. 2019a).

Hormones play a crucial role in the development of cereal endosperm. Indole-3-acetic acid (IAA) is synthesized and accumulated in fertilized maize kernels and rice grains, where it promotes endosperm development (Basunia and Nonhebel 2019). Cytokinin (CK) positively regulates endosperm cell division rate and enhances starch accumulation (Zhang et al. 2020). Abscisic acid (ABA) is involved in the cellularization of endosperm (Sreenivasulu et al. 2010), while brassinolide (BR) plays a role in endosperm development and starch accumulation (Zhang et al. 2020). The key genes involved in hormone biosynthesis, such as *OsYUCs*, *OsTAR1* and *ZmEHD1* (for IAA), *SEG8* and *DG1* (for ABA), and *DWARF4* (for BR), are important regulators for starch content (Abu-Zaitoon et al. 2012; Qin et al. 2021; Sreenivasulu et al. 2010; Wang et al. 2020; Zhang et al. 2020). In addition, epigenetic mechanisms, such as DNA methylation and histone modification, also play a crucial role in regulating cereal endosperm development (Zhao and Zhou 2012 and Zhang et al. 2018).

In recent years, the development of high-throughput sequencing technology and the release of numerous plant reference genomes have made it easier to develop high-density molecular markers covering the whole genome of plant species. Moreover, the interactive utilization of global germplasm resources has provided convenience for the construction of large-scale genetic populations. Consequently, genome-wide association study (GWAS) has become a powerful tool for dissecting quantitative traits (Yamaguchi-Kabata et al. 2008). For example, over the past decade, the most widely used maize association mapping panel (AMP), which represents global maize diversity collected from tropical/subtropical and temperate germplasms, was constructed by Yang et al. (~527 inbred lines) for GWAS (Yang et al. 2011). Using this population, numerous QTL and candidate genes for corresponding traits have been cloned and proposed (Li et al. 2013; Sun et al. 2022; Yang et al. 2013; Zhang et al. 2021). However, GWAS has limited power to detect minor alleles due to the requirement of a minimum allele frequency (MAF) higher than 0.05 (Everett et al. 2020), leading to the exclusion of small-effect genes. As we all know, increasing the density of molecular markers is necessary to improve the detection efficiency of GWAS. Zhang et al. (2016b) conducted a GWAS on drought-related metabolic changes using an enlarged SNP panel (156,599 SNPs) and detected 63 significant QTL, including 56 novel loci compared to a previous study (Setter et al. 2011). Similarly, using an enlarged SNP panel obtained by identity by descent (IBD) and *k*-nearest neighbor (KNN) imputation methods, GWAS was conducted on 17 agronomic traits of 513 maize inbred lines, revealing numerous significant loci, including known and some novel QTL (Yang et al. 2014). Except for enlarged genotype data, innovation in statistical models has greatly contributed to improving the detection efficiency of GWAS. For example, the GCIM-QEI model can detect QTL-by-environment interaction loci, and the IIIVmrMLM model can detect dominant effect loci (Li et al. 2022; Zhou et al. 2022). Therefore, taking into account both enlarged genotype data and efficient statistical models is crucial for GWAS.

Understanding the genetic architecture of starch content is crucial for identifying candidate genes. Using a recombinant inbred line (RIL) population with CI7 and K22 as parental lines. Six QTL that affect starch content in maize kernels were identified. Each of these QTL explained 4.07 to 10.6% of phenotypic variation. Furthermore, seven genes were considered as potential causal genes, with four of them acting as regulators of starch biosynthesis (Wang et al. 2015). Another study identified 13 QTL using four double haploid (DH) populations, with 12 genes located within these QTL being implicated in starch synthesis (Zhang et al. 2022). In another research, employing single linkage mapping, joint linkage

mapping, and a genome-wide association study, 50 QTL were identified by using a multi-parent population, of which 18 were novel. Notably, *ZmTPS9* was identified as the causal gene, encoding a trehalose-6-phosphate synthase. Knocking out *ZmTPS9* resulted in increased starch content and grain weight in maize (Hu et al. 2021). These findings broaden our understanding of the genetic basis for starch content. Nevertheless, starch content is a quantitative trait, and further research is essential to unveil its genetic and molecular mechanisms.

In this study, we re-analyzed the genetic basis of starch content for 261 maize inbred lines using an enlarged SNP panel and improved statistical models; the main purpose was (i) to identify novel loci that may regulate maize starch content, (ii) to filter not yet reported candidate genes, and (iii) to develop markers for use in marker-assisted selection of maize kernels with starch content. Our research aims to provide new insights into improving maize kernel starch content.

Materials and methods

Phenotype resources

Phenotypic data for this study was obtained from 261 maize inbred lines cultivated across grown in three different environments, each with three replicates. These 261 inbred lines were randomly selected from a panel of 513 inbred lines used for association mapping. Among them, 71 inbred lines originated from tropical/subtropical regions, while 190 were from temperate regions. All these inbred lines were planted in Ledong, Hainan Province (latitude 18.75° N, longitude 109.17° E), for the years 2011, 2012, and 2013. The region experiences an annual average precipitation of 1181 mm, an annual average temperature of 23 °C, an annual average sunshine duration of 1039.6 h, and an accumulated temperature of 9300.7 °C (Liu et al. 2016a). The starch content of each line was measured with three repetitions, and the best linear unbiased prediction (BLUP) was calculated for the combined data from all three environments and replicates. BLUP value was used as the phenotypic data for GWAS in this study.

Genotype resources

In this study, by combining the MaizeSNP50 BeadChip and RNA sequencing, an enlarged SNP panel containing 558,629 high-quality SNPs (B73_RefGen_v2, referred 0.56M) using two-step approaches, identited by descent (IBD) and the *k*-nearest neighbor (KNN) algorithm, was obtained (Yang et al. 2014). The set of genotype data covers the entire maize genome with a minimum allele frequency (MAF) of at least 0.05 (MAF≥0.05) and can be downloaded from the Maizego website (http://www.maizego.org/Resources.html).

Statistical model

To control for both type I (false positive) and type II (false negative) error rates, three models were compared: generalized linear model (GLM) with population structure as a fixed effect (GLM + Q), mixed linear model (MLM) with relative kinship as a random effect (MLM + K), and MLM with both population structure and kinship as fixed and random effects (MLM + Q + K), respectively. Specifically, the GLM can be represented as $y = X\alpha + Z\beta + e$, while MLM can be represented as $y = X\alpha + Z\beta + W\mu + e$. Here, y is the trait value, $X\alpha$ represents the population structure or Q matrix as a fixed effect, $Z\beta$ represents SNP or marker effect as a fixed effect, $W\mu$ represents the kinship matrix as a random effect, and e represents the residual (Yu et al. 2006). It is essential to provide detailed information regarding the Q matrix as the Q model is the most suitable for this research. The number of subgroups (K) was set from 1 to 15 and was used to identify the optimal K value; this was achieved by conducting 150,000 MCMC (Markov chain Monte Carlo) replications and 100,000 burn-ins in both STRUCTURE and INSTRUCT software. These tools were employed to estimate population structure and create subpopulations. In the STRUCTURE software, combining the log-likelihood of data (LnP(D)) and an ad hoc statistic ΔK determines the most suitable K value. Meanwhile, in INSTRUCT, LnP(D) and deviance information criterion were used to define the optimal K. To consolidate the results obtained from replicate simulations conducted in

STRUCTURE and INSTRUCT, CLUMPP software was used. Inbred lines with probabilities greater than or equal to 0.60 were assigned to their respective subpopulations, while lines with probabilities less than 0.60 were grouped into a mixed category (Yang et al. 2011). To evaluate the performance of the three models, quantile-quantile (QQ) plots were generated for each model using the best linear unbiased prediction (BLUP) of starch content. An optimal model was determined based on a QQ plot that had a line close to 1:1 with a distinct tail that deviated upwards, indicating that well-controlled type I and type II errors and a true association with causal polymorphism(s) (Zhang et al. 2010).

Genome-wide association analyses

The genome-wide association study (GWAS) was performed using TASSEL 3.0 software (Bradbury et al. 2007). To account for the linkage disequilibrium among SNP markers, the effective number of markers (En) was calculated using GEC software (Yang et al. 2010) as 250,345 for 0.56M (558,629 SNPs) SNPs, respectively. Additionally, to avoid false negative (type II error) and be able to detect more small effect loci, an appropriately adjusted threshold of 2.07×10^{-5} was used for the 0.56M SNPs, which is commonly used in plant genome-wide association study.

Candidate gene analyses

To identify potential candidate gene associated with starch content, we defined significant QTL by using previously estimated linkage disequilibrium (LD) distance, and a 100-kb QTL interval was defined for 0.56M SNPs, with 50 kb upstream and downstream of each significant SNP (Yang et al. 2014). The candidate gene was identified using the filtered working gene list from the B73 reference genome (RefGen_v2) obtained from MaizeGDB. The candidate gene was annotated using InterProScan (http://www.ebi.ac.uk/interpro/scan.html), and expression patterns in maize organs were analyzed to predict the potential relationship with starch content (Hoopes et al. 2019). The most likely candidate gene within each QTL was selected based on its annotation or contained the peak SNP. If there were no genes within the interval, the neighboring gene of the peak SNP was considered the most likely candidate gene. The rule of QTL naming is as follows: q + trait + serial number, for example, [*qSc1*, *qSc* (Starch content) *1* (serial number)]. Additionally, using the "lm" package in R software, we employed a multiple linear regression model to assess the total phenotypic variation accounted for all QTL (Zhang et al. 2016b).

Gene Ontology (GO) enrichment analyses

To identify enriched Gene Ontology (GO) terms, we performed GO enrichment analysis using OmicShare Tools (https://www.omicshare.com/tools) (Ding et al. 2019). The analysis involved mapping genes expressed in maize endosperm to various sets in the GO database (http://www.geneontology.org/). The number of genes in each set was counted, and a list of genes with a specific GO function and the number of genes in each function was obtained. The top 30 GO terms with the minimum *P* values were selected for analysis and visualization.

Linkage disequilibrium analyses

The extent of linkage disequilibrium (LD) was estimated using the squared correlation of paired SNPs, which was computed using the "genetics" package in R (version 4.1.1). An LD plot was then generated with the "LDheatmap" package in R.

Haplotype analyses of ZmAPC4

BLUP value of starch content with 261 inbred lines was used as phenotype data. All SNPs located in *ZmAPC4* were used as genotype data, and then combing them for haplotype analyses. According to the number of inbred lines that carry different haplotypes with starch content from high to low, they were named hap1, hap2, hap3, and hap4, respectively. To ensure robustness in our analysis, haplotypes consisting of fewer than 10 inbred lines were excluded from further consideration. Additionally, taking into account the peak single nucleotide polymorphism (SNP) chr4.S_175584318, which is associated with *ZmAPC4*, two distinct haplotypes carrying either the G allele or the T allele were selected for comprehensive haplotype analysis. Significant differences between different haplotypes were determined using Student's *t*-test.

Construction of phylogenetic tree

To bolster the credibility of *ZmAPC4*, we searched all genes encoding the WD40 domain that have been previously documented in *Arabidopsis*, maize, and rice. Using the amino acid sequence of these genes, we conducted an amino acid sequence alignment with the neighbor-joining (NJ) method (Saitou and Nei 1987) within the MEGA X software. Subsequently, the resulting phylogenetic tree was annotated using the iTOL online tool (https://itol.embl.de/).

Development of molecular markers

For the peak SNP chr4.S_175584318, which exhibits two alleles (GG or TT), we developed molecular markers to distinguish inbred lines based on their starch content. To achieve this, we selected inbred lines with higher starch content (carrying the GG allele) and lower starch content (carrying the TT allele), respectively. To design the molecular markers, we utilized primer 1 to amplify a 504bp fragment that includes the peak SNP. We employed the dCAPS Finder 2.0 program available at http://helix.wustl.edu/dcaps/ (Neff et al. 2002) to develop dCAPS markers and design nearly matched primers, referred to as primer 2. Primer 2 was specifically designed to amplify a 251bp fragment that contains the NdeI restriction site ('CATATG') from the 504bp fragment. Following the amplification with primer 2, the resulting 251bp fragment was extracted and purified using a 4% agarose gel. Subsequently, the purified product underwent digestion with NdeI endonuclease (New England Biolabs R0111V). Detailed information regarding the primers, PCR system, and enzyme digestion can be found in Table S3.

Results

Optimized model and expanded genotype

The characterization of starch content in the AMP using a near-infrared analyzer (NIA) is an important step for breeding high-quality maize varieties. In a previous study, the starch content of 261 maize inbred lines was measured by NIA, a genome-wide association study (GWAS) using a mixed linear model (MLM) with principal components (PCs) and kinship (K) as a model (PCs + K) was performed. However, this model was found to be too stringent in reducing false positive (type I error) compared to other models (Figure S1 and Figure S2) (Liu et al. 2016a). To improve the accuracy of the results, we used three different models (Q, K, and Q+K) for GWAS, where the Q model only accounts for population structure, the K model only accounts for kinship, and the Q+K accounts for both population structure and kinship. To test whether increasing marker density could further improve GWAS detection power for starch content, we expanded the genotype to 558,629 high-quality SNPs and repeated the analysis using the Q, K, and Q+K models. However, both the K and Q+K models still showed too much false negative, while

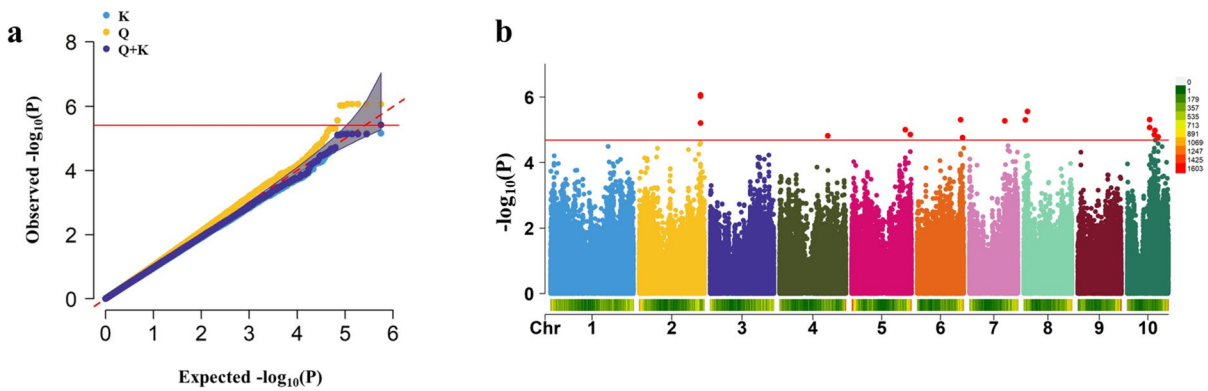

Fig. 1 **a** QQ plot of Q, K, and Q+K models for starch content based on 0.56M SNPs. The red line represents the significant threshold was 3.99×10^{-6} (1/250345). **b** Manhattan plot of Q model based on 0.56M SNPs. The red line represents the significant threshold was 2.07×10^{-5} (1/48393)

Table 1 All genes within significant QTL associated with starch content by GWAS with an enlarged SNP panel

ID[a]	Chr	Peak SNP	*P* value[b]	R^2 (%)[c]	Gene	Function annotation
1	2	chr2.S_223035929	8.51E−07	9.61	*GRMZM2G056335*	UDP-glucosyl transferase
2	2	chr2.S_223035948	8.51E−07	9.61	*GRMZM2G056335*	UDP-glucosyl transferase
3	2	chr2.S_223035985	8.51E−07	9.61	*GRMZM2G056335*	UDP-glucosyl transferase
4	2	chr2.S_223036014	8.51E−07	9.61	*GRMZM2G056335*	UDP-glucosyl transferase
5	2	chr2.S_223036020	8.51E−07	9.61	*GRMZM2G056335*	UDP-glucosyl transferase
6	2	chr2.S_223036032	9.35E−07	9.62	*GRMZM2G056335*	UDP-glucosyl transferase
7	2	chr2.S_223036071	9.35E−07	9.62	*GRMZM2G056335*	UDP-glucosyl transferase
8	4	chr4.S_175584318	1.52E−05	7.60	*GRMZM2G053766*	WD40 repeat-like superfamily protein
9	5	chr5.S_195128171	9.94E−06	8.18	*GRMZM2G468585*	NA
					AC194912.3_FG003	NA
10	5	chr5.S_213004279	1.40E−05	8.03	*GRMZM2G080930*	Serine/arginine
					GRMZM2G005791	Cyclic nucleotide-gated channel
					GRMZM2G080843	Transmembrane amino acid transporter family protein
11	6	chr6.S_157907570	4.92E−06	8.42	*GRMZM2G138067*	NA
					GRMZM2G138165	NA
					GRMZM2G138076	Outer membrane OMP85 family protein
					GRMZM2G032847	Mitochondrially targeted single-stranded DNA binding protein
					GRMZM2G049091	Transcription initiation factor IIF beta subunit
					GRMZM2G033406	NA
12	6	SYN19442	1.74E−05	7.39	*GRMZM2G023207*	NA
					GRMZM2G022619	Ribosomal protein L18ae/LX family protein
					GRMZM2G023190	F-box/RNI-like superfamily protein
					GRMZM2G022453	Ribosomal protein L18ae family
					GRMZM2G317596	AP2-EREBP-transcription factor
					GRMZM2G154626	F-box/RNI-like/FBD-like domains-containing protein
					GRMZM2G023133	Cytochrome b561/ferric reductase transmembrane protein family
					GRMZM2G409881	WOX9C protein
13	7	chr7.S_128433541	5.34E−06	8.11	*AC214483.3_FG002*	NA
					GRMZM2G314679	SNARE-like superfamily protein
					GRMZM2G021049	SAUR-like auxin-responsive protein family
14	8	chr8.S_6150835	4.99E−06	8.60	*GRMZM2G044819*	HAD-superfamily hydrolase subfamily IG 5\'-nucleotidase
15	8	chr8.S_13913478	2.75E−06	9.04	*GRMZM2G114486*	NA
					GRMZM5G894420	NA
					GRMZM2G020484	Peptide transporter
16	10	chr10.S_82049872	8.50E−06	7.45	*GRMZM2G358161*	NOD26-like intrinsic protein
					GRMZM5G828229	Monodehydroascorbate reductase
					GRMZM5G856084	DNAJ heat shock family protein
					GRMZM5G832300	RING/U-box superfamily protein
17	10	PZE-110043085	4.88E−06	8.27	*AC188036.3_FG003*	NA
					GRMZM2G032684	Oxidoreductase family protein
					GRMZM2G149786	LSD1-like
					GRMZM2G042980	NA
18	10	chr10.S_98706750	1.42E−05	7.09	*GRMZM2G007721*	UDP-Glycosyltransferase superfamily protein

Table 1 (continued)

ID[a]	Chr	Peak SNP	*P* value[b]	R^2 (%)[c]	Gene	Function annotation
19	10	PZE-110053448	1.05E−05	7.31	*GRMZM2G347693*	NA
20	10	chr10.S_112584896	1.65E−05	7.02	*GRMZM2G021802*	DNA repair metallo-beta-lactamase family protein
					GRMZM2G022180	Calnexin homolog2
					GRMZM5G837325	NA
					GRMZM2G021725	High-affinity K+ transporter
21	10	chr10.S_112585868	1.65E−05	7.02	*GRMZM2G021802*	DNA repair metallo-beta-lactamase family protein
					GRMZM2G022180	Calnexin homolog2
					GRMZM5G837325	NA
					GRMZM2G021725	High-affinity K+ transporter

[a]The ID of the loci identified by a GWAS

[b]*P* value of the corresponding trait calculated by optimal model

[c]The phenotypic variance explained by the corresponding locus

the Q model consistently outperformed the other two best among the three models (Fig. 1a). In conclusion, the Q model is the most appropriate choice for this research. Additionally, our study also suggested that increasing marker density can improve the statistical power of GWAS and more SNPs/loci were detected, and the choice of the appropriate model is crucial for successful GWAS.

GWAS

The Q model was selected and used to interpret GWAS results of starch content. A total of 21 significant SNPs were identified (Fig. 1b, and Table 1), indicating that expanding the marker density and using appropriate thresholds can improve the detection power of GWAS (only four loci were detected in the previous study) (Liu et al. 2016a). To further understand the genetic basis underlying starch content, the 21 significant SNPs were categorized into 14 QTL based on the definition of QTL. Each QTL could explain the phenotypic variation (R^2) ranging from 7.02 to 9.62%, with an average of 8.42% and the total phenotypic variation explained by all QTL is 47.66%. These QTL are likely to be associated with starch content and defined as starch-content candidate loci. Furthermore, at chr.2, *qSc1*(222.99 Mb-223.09 Mb) had a powerful ability to explain 9.61% of the phenotypic variation for starch content (Table S1). Remarkably, seven significant SNPs co-located in *qSc1*, indicating that *qSc1* is a crucial region. Furthermore, only one gene (*GRMZM2G056335*) was identified within *qSc1*, and its annotation as UDP-glucosyl transferase has been reported to be associated with heat-stress-induced leaf senescence (Han et al. 2023). Apart from its potential impact on starch content, it also appears to have a role in conferring resistance to abiotic stress. These findings strongly suggested that *qSc1* represents a genetic hotspot region. In conclusion, these results provide valuable insights into the accumulation of useful information concerning starch content in maize kernels. Analyzing the candidate gene responsible for underlying the QTL could reveal even more insights into the genetic basis of starch content.

Candidate gene analysis about multiple loci

After analyzing the GWAS results, we identified a total of 42 genes involved in various functions, such as transcription factors (e.g., *GRMZM2G138165*), enzymes (e.g., *GRMZM2G056335*), and proteins (e.g., *GRMZM2G138076*, *GRMZM2G005791*) (Table 1). As the maize endosperm is the main organ rich in starch and accounts for more than 90% of the kernel dry weight, we found about half of all genes (22/42) expressed in the maize endosperm; it suggested their potential role in regulating starch content. Only two genes have been reported to be directly involved in endosperm development. One of them is *GRMZM2G080843*, which plays a regulatory role in starch biosynthesis in maize endosperm (Finegan et al. 2022). Another is *GRMZM2G022453*, which is

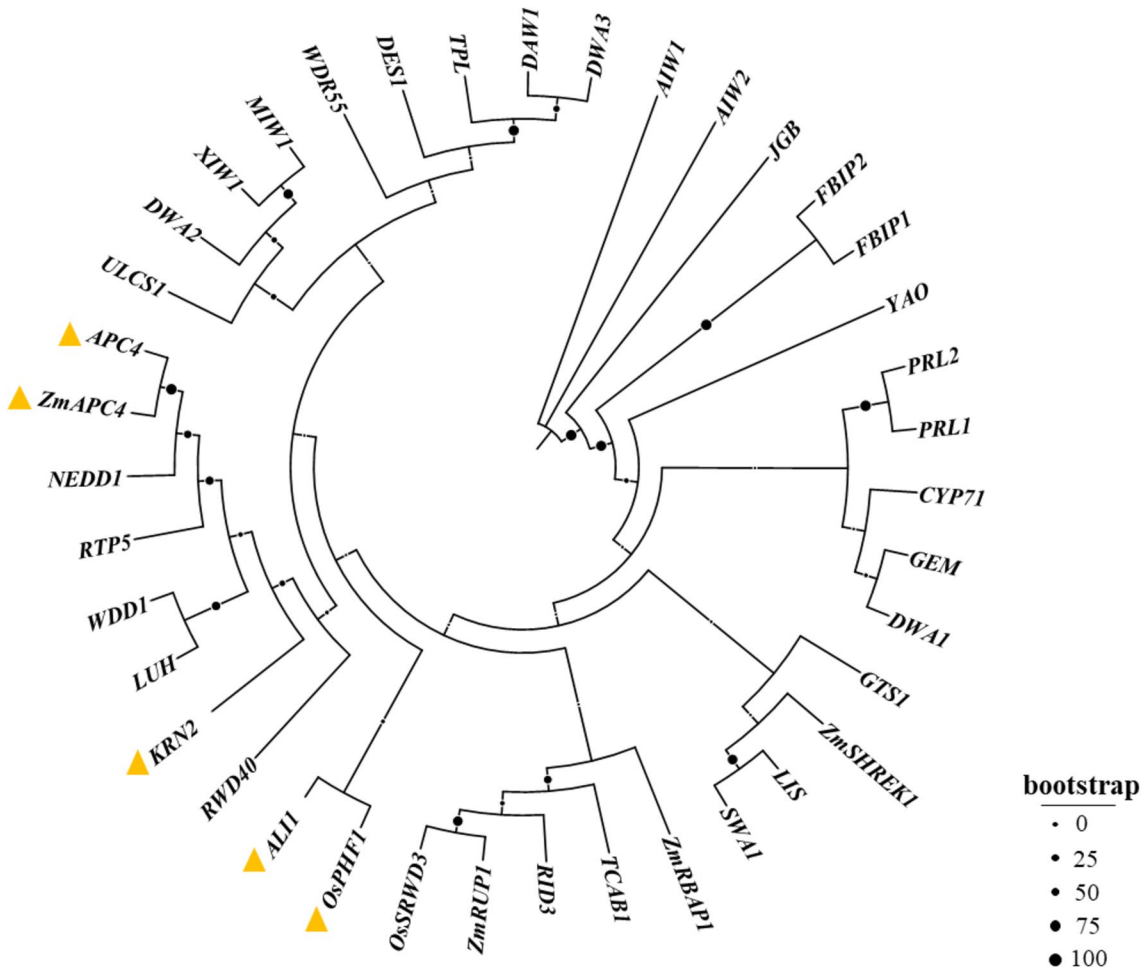

Fig. 2 The phylogenetic tree of *ZmAPC4* and other genes that encode WD40 protein domain

also implicated in endosperm development and has an indirect impact on starch synthesis (Song et al. 2021). Overall, it's noteworthy that only two genes had been characterized in terms of their roles in starch synthesis within the endosperm, the remaining forty genes represent novel candidates, highlighting the potential value of using an expanded SNP panel to gather additional insights into the genetic basis of starch content in maize kernels. Notably, *GRMZM2G053766* has a high expression level during endosperm development, and it encodes a WD40 family protein and shares homology with anaphase-promoting complex 4 (*APC4*) in *Arabidopsis*. Mutations in *apc4* have been linked to abnormal endosperm development and cell cycle disorders (Guo et al. 2018). Another evidence comes from *TRANSPARENT TESTA GLABRA1* (*TTG1*), which also encodes a WD40 repeat transcription factor, it has been demonstrated to play a role in seed storage accumulation in *Arabidopsis*. In *ttg1-1* mutant seeds, there is a significant increase in dry weight, primarily attributed to elevated starch content, total protein, and fatty acids (Chen et al. 2015a). Given these findings, we named *GRMZM2G053766* as *ZmAPC4*. Subsequently, we filtered genes encoding the WD40 domain that had been previously reported in *Arabidopsis*, maize, and rice. Using the amino acid sequence of these genes, we constructed the phylogenetic tree of *ZmAPC4* (Fig. 2) employing the neighbor-joining (NJ) method. The results revealed that *ZmAPC4* shares the highest homology with *APC4* in *Arabidopsis*. Additionally, we also observed homology between *ZmAPC4* and the

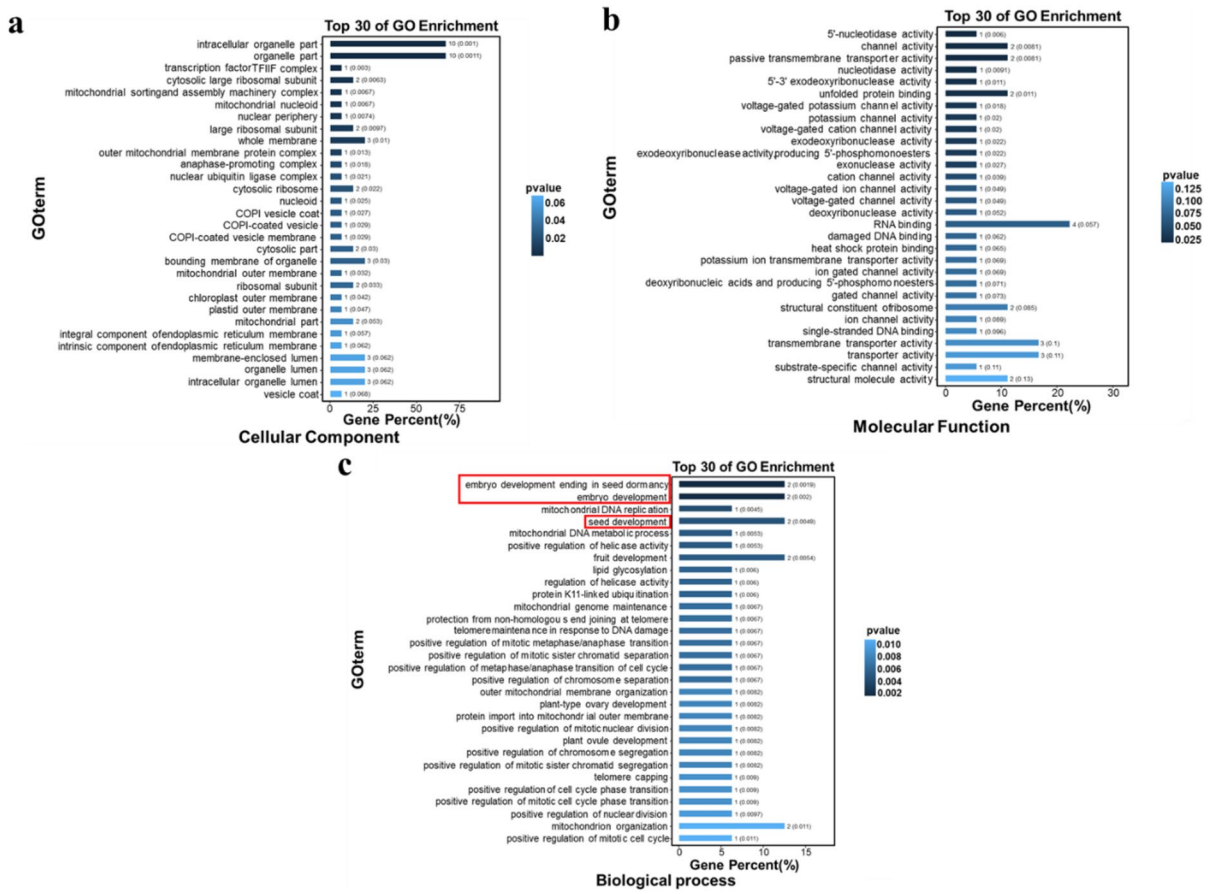

Fig. 3 Top 30 of GO enrichment entries by using 22 genes expressed in maize endosperm

reported genes *KRN2* (Chen et al. 2022), *ALI1* (Best et al. 2021) in maize, and *OsPHF1* in rice (Chen et al. 2015b). In conclusion, the homology information strengthens the credibility of *ZmAPC4* as a candidate gene that may influence starch content.

Gene Ontology enrichment analysis

We conducted a GO enrichment analysis on the 22 genes that were expressed in the maize endosperm and found significant enrichment in several categories, including intracellular organelle part (cellular component), transmembrane transporter activity, RNA binding (molecular function), and embryo development (biological process) (Fig. 3). Importantly, *ZmAPC4* was enriched in 18 of the top 30 GO terms and showed significant enrichment in biological processes related to embryo and seed or fruit development (GO:0009793, GO:0009790, and GO:0048316) (Table S2). These findings suggest that *ZmAPC4* plays a critical role in maize kernel development and may be involved in regulating starch synthesis.

Haplotype analysis of ZmAPC4

To analyze the haplotype of *ZmAPC4*, we extracted all SNPs within one LD decay distance (±100kb) upstream and downstream of the peak SNP (chr4.S_175584318, $P = 1.52E{-}05$) (Fig. 4a), we found strong linkage relationship between other SNPs and peak SNP (Fig. 4b). chr4.S_175584318 (TT/GG) is located in the 3′ UTR region of *ZmAPC4* and does not result in a change of encoded amino acid, but allele variation can alter mRNA stability and lead to changes in expression (Pal et al. 2011). Based on 0.56M SNPs, we filtered all SNPs located in *ZmAPC4* as genotype

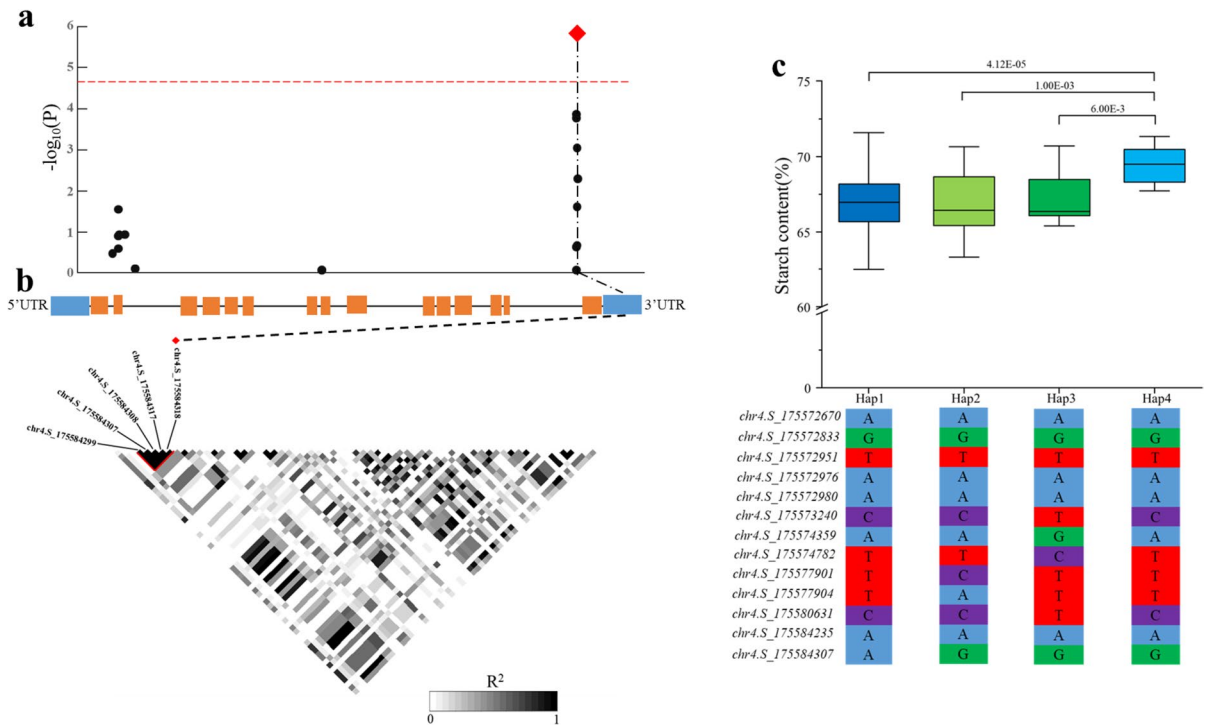

Fig. 4 *ZmAPC4* affected starch content in maize kernels. **a** Enlarged Manhattan plot of the lead SNP; red diamonds represent the lead SNP. The red dotted line represents the threshold $-\log10(P) \geq 4.68$ ($P \leq 2.07 \times 10^{-5}$). **b** R^2 values of SNPs associated with *ZmAPC4*; the lead SNP located at 64 bp downstream within 3'UTR of *ZmAPC4*. **c** Using one-way ANOVA to perform haplotype analysis of *ZmAPC4*, considering the uneven number of inbred lines among haplotypes, the Games-Howell method was employed for pairwise comparisons. The maize inbreds which carrying hap4 had highest starch content

data. BLUP value of starch content with 261 inbred lines was used as phenotype data. Combing genotype and phenotype for haplotype analysis after removing the missing SNPs. It was observed that there are significant differences in starch content between hap1 and hap4, hap2 and hap4, as well as hap3 and hap4 (Fig. 4c), The average difference in starch content between hap1, hap2, hap3, and hap4 is approximately 2.6%. Among these haplotypes, hap4 (AGTAACATT TCAG) consisted of 11 inbred lines that exhibited the highest starch content. This suggested that these particular inbred lines may harbor favorable allele variants associated with increased starch content (Table 2). Regarding the peak SNP chr4.S_175584318, significant differences were observed between the two haplotypes carrying the T or G (Fig. 5a). Based on the identification of these favorable genomic regions, molecular marker-assisted selection could be employed to differentiate the starch content of various maize germplasms and enhance the starch levels in modern maize breeding programs.

Development molecular markers of ZmAPC4

Molecular marker-assisted selection is a valuable approach that complements traditional breeding methods, enhancing efficiency in the breeding process (Guo et al. 2021). In our study, we successfully developed dCAPS markers capable of categorizing inbred lines based on their starch content (Table S3). To implement the markers, DNA fragments containing the peak SNP chr4.S_175584318 were subjected to digestion using the NdeI enzyme; subsequently, the resulting fragments were separated using 4% agarose gel electrophoresis. Notably, distinct banding patterns were observed between the eight high-starch lines (carrying the GG allele) and the eight low-starch lines (carrying the TT allele) (Table S4 and Fig. 5b). These findings yield two important implications. Firstly, *ZmAPC4* emerges as a stable candidate gene associated with starch content. Secondly, the developed molecular markers can be effectively utilized to screen other maize varieties with either higher or lower starch content. Consequently,

Table 2 Comparison of haplotypes for starch content using BLUP value of 261 inbred lines

Haplotypes	chr4.S _175572670	chr4.S _175572833	chr4.S _175572951	chr4.S _175572976	chr4.S _175572980	chr4.S _175573240	chr4.S _175574359	chr4.S _175574782	chr4.S _175577901	chr4.S _175577904	chr4.S _175580631	chr4.S _175584235	chr4.S _175584307	Inbred lines	starch content
Hap1	A	G	T	A	A	C	A	T	T	T	C	A	A	174	66.83 ± 1.96
Hap2	A	G	T	A	A	C	A	T	C	A	C	A	G	15	66.74 ± 1.90
Hap3	A	G	T	A	A	T	G	C	T	T	T	A	G	14	66.89 ± 2.05
Hap4	A	G	T	A	A	C	A	T	T	T	C	A	G	11	69.41 ± 1.09

our research significantly contributes to improving breeding efficiency and provides a valuable reference for the development of new maize varieties characterized by high starch content.

Discussion

Starch is a major component of plant endosperm and is mainly composed of two types: amylose and amylopectin (Jeon et al. 2010). In our study, we identified two genes, *GRMZM2G056335* and *GRMZM2G007721*, which encode UDP-glucosyl transferase (UDPG). UDPG-related genes, such as *du* and *waxy* in rice, have been found to affect amylose content in grains (Kaushik and Khush 1991 and Zhang et al. 2019b), while *flos* (*flo1* and *flo2*) affects the transparency of endosperm and produces a starchy endosperm (She et al. 2010). Therefore, our findings suggest that *GRMZM2G056335* and *GRMZM2G007721* may be the potential targets for affecting starch content in maize kernels.

Starch biosynthesis in cereals is regulated by various transcription factors (TFs), including *MYC*, *EREBP*, *bHLH*, and *PPR* family transporters in rice (Bello et al. 2019; Wu et al. 2020; Zhu et al. 2003), *MYBs* and *DOFs* in maize (Wu et al. 2019; Xiao et al. 2017), and *AP2/EREBP* and *bZIP* in wheat (Liu et al. 2016b; Song et al. 2020). Notably, our study identified *GRMZM2G317596*, which encodes an *AP2/EREBP* transcription factor (Table 1). In rice, the homolog gene *RSR1* was found to regulate starch biosynthesis, and its mutant led to the up-regulation of genes involved in starch synthesis, an increase in amylose content, and changes in amylopectin structure, which altered the morphology of starch grains (Fu and Xue 2010). Therefore, *GRMZM2G317596* could be regarded as a potential candidate gene affecting starch content in maize. In addition to TFs, several long non-coding RNAs (lncRNAs) have also been found to play a role in starch biosynthesis. Overexpression of *lncRNA_2308*, *lncRNA_1267*, and *lncRNA_1631* reduced the expression of *GBSSI*, resulting in a decrease in starch content and grain weight in rice (Zheng et al. 2019). However, the complexity of transcriptional regulation in starch biosynthesis implies that mutation in a single TF gene may not result in significant changes in endosperm development and

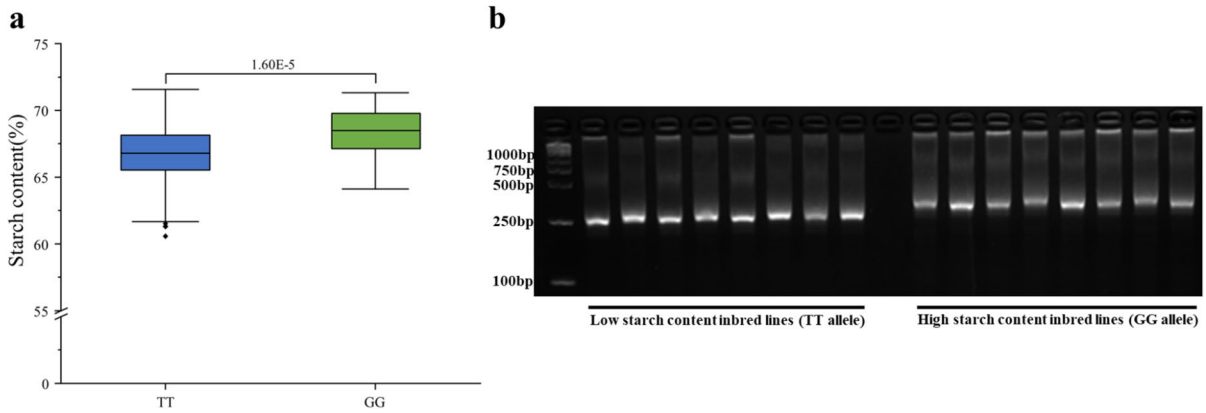

Fig. 5 Difference between maize inbred lines which have TT or GG allele for peak SNP. **a** Difference analysis of maize inbreds have T or G haplotype. **b** The bands of eight inbred lines have TT allele and eight inbred lines have GG allele after using agarose gel electrophoresis

starch content. The complexity suggests that the endosperm has established feedback mechanisms to respond to internal and external changes. Therefore, co-expression analysis and genetic analysis are powerful tools for identifying candidate genes that regulate endosperm development and starch content. In particular, GWAS has become an efficient method for detecting and analyzing the genetic mechanisms of quantitative traits.

To ensure the accuracy of GWAS results, it is important to consider the trait sensitivity and choose the appropriate model with high statistical power and low error rate (Chang et al. 2018). In this study, we employed an enlarged genotype dataset and applied three different statistical models (Q, K, and Q+K) to conduct GWAS. After a comprehensive comparison of the results, we found that the K and Q+K models had a more stringent control on false positive. On the other hand, the Q model had the best control effect on false positive, as indicated by the QQ plot, and was found to be a suitable model for our research.

In a previous GWAS for kernel starch content, only four SNP-trait associations underlying four candidate loci were identified using a set of 52,370 SNPs (Liu et al. 2016a). In the present study, an enlarged panel of high-density SNP panel (558,629) was obtained from RNA sequencing data performed on 368 of the 513 lines used in the previous study with a minor allelic frequency greater than 0.05 (Yang et al. 2014). This enabled the identification of 14 new loci significantly associated with maize kernel starch content, as well as 21 additional significantly associated SNPs that were not detected in the previous study with smaller density markers (Table S1). Interestingly, the four loci significantly associated with starch content in the previous study were not significant in the current study. This was likely due to the different *P*-values of the same SNPs in the present study, which did not meet the suggestive threshold ($P \leq 0.05/48,277$) in the expanded GWAS (McGeachie et al. 2015). The increase in significant SNPs in the present study was primarily due to the higher marker density, which increased statistical power and enabled the identification of minor effects and unbalanced allele frequency loci (Gong et al. 2013; Tedja et al. 2018).

In the current study, we found 21 significant SNPs associated with maize kernel starch content, involving 14 novel QTL containing 42 genes. Out of these, 29 genes had functional annotation. Some of the QTL we identified had been reported in previous studies. For example, *qSc6* was located on chromosome 6 within the interval of 164.64 Mb-164.74 Mb (Table S1), which was previously identified as a QTL for starch content using BLUP value with epistatic QTL by a GWAS (Hu et al. 2021). Based on these findings, we suggested that *qSc6* may be considered as a stable QTL that regulates starch content. Eight genes were within *qSc6*, including those encoding ribosomal family protein (*GRMZM2G022453*, *GRMZM2G022619*), F-box family protein (*GRMZM2G154626*, *GRMZM2G023190*), and the *AP2-EREBP* transcription factor (*GRMZM2G317596*). In wheat and rice, *AP2/EREBP* transcription factors have been reported to be closely related to starch content (Fu and Xue

2010; Liu et al. 2016b). This suggests that these genes may be conserved during evolution and could have similar functions in maize. We also identified *qSc2*, which is located on chromosome 4. The significant SNP of *qSc2* (chr4.S_175584318, $P = 1.52 \times 10^{-5}$) was found to be co-located with chr4.S_165621095 (the distance between two SNPs was less than 10Mb) in a previous study (Li et al. 2018b). Only one gene, *ZmAPC4*, was located within *qSc2*; it encodes a WD40 repeat-like superfamily protein. *APC4* has been demonstrated to influence endosperm development, and WD40 proteins have been shown to impact starch accumulation in *Arabidopsis* seeds (Chen et al. 2015a; Guo et al. 2018). GO analysis confirmed that *ZmAPC4* significantly affects the progression of grain development (Fig. 3), and the haplotype analysis revealed a significant difference in starch content between maize inbred lines carrying hap4 compared to those carrying other haplotypes (Fig. 4c). Furthermore, the peak SNP chr4.S_175584318, which is associated with the *ZmAPC4*, displayed an interesting pattern where inbred lines carrying the GG allele exhibited higher starch content compared to those carrying the TT allele (Fig. 5a). Capitalizing on the information from the peak SNP, we successfully developed dCAP markers capable of distinguishing between maize-inbred lines with higher or lower starch content. These molecular markers can be widely applied to assess and differentiate the starch content of various maize lines (Fig. 5b), This advancement could significantly enhance breeding efficiency and offer a valuable tool for developing new maize varieties with high starch content. These findings suggest that *ZmAPC4* plays a role in regulating starch content. Genetically modified (GM) technology such as gene silencing, knockout, and overexpression could be employed to further verify the functions of these genes in different cereal crops, such as wheat and rice.

Conclusion

Overall, our study employed an expanded SNP panel and a more suitable statistical model to re-analyze the published data on maize kernel starch content, resulting in the identification of several novel genetic loci through GWAS. We also predicted potential candidate genes that may regulate starch content, which could be useful for improving the efficiency of maize breeding through the development of molecular markers. Our findings provide a valuable reference for enhancing grain yield and could contribute to the development of more productive and sustainable agricultural practices.

Acknowledgements We thank Jianbing Yan's group at Huazhong Agricultural University for providing maize materials and genotype data.

Author contribution X. Z. designed the study. X. Z. and J. T. supervised the study. H. D., J. L., L. S., X. X., S. X., Y. S., X. J., Z. X., J. G., Y. W., H. X., and DD performed the experiment and analyzed the data. HD and XZ prepared the manuscript and all authors read and approved the manuscript.

Funding This research was supported by the National Natural Science Foundation of China (32171980), a project funded by the China Postdoctoral Science Foundation (2020M682295), the Henan Province Science and Technology Attack Project (232102110181), a first-class postdoctoral research grant in Henan Province (202001032), the Henan Provincial Higher Education Key Research Project (24B210003), and the Research Start-up Fund for Youth Talents of Henan Agricultural University (30500563).

Data availability The genotype dataset included in this study is available in an online repository http://www.maizego.org/Resources.html.

Declarations

Conflict of interest The authors declare no competing interests.

References

Abu-Zaitoon YM, Bennett K, Normanly J, Nonhebel HM (2012) A large increase in IAA during development of rice grains correlates with the expression of tryptophan aminotransferase OsTAR1 and a grain-specific YUCCA. Physiol Plant 146(4):487–499. https://doi.org/10.1111/j.1399-3054.2012.01649

Basunia MA, Nonhebel HM (2019) Hormonal regulation of cereal endosperm development with a focus on rice (Oryza sativa). Funct Plant Biol 46(6):493–506. https://doi.org/10.1071/FP18323

Bello BK, Hou Y, Zhao J, Jiao G, Wu Y, Li Z, Wang Y, Tong X, Wang W, Yuan W, Wei X, Zhang J (2019) NF-YB1-YC12-bHLH144 complex directly activates Wx to regulate grain quality in rice (Oryza sativa L.). Plant Biotechnol J 17(7):1222–1235. https://doi.org/10.1111/pbi.13048

Best NB, Addo-Quaye C, Kim BS, Weil CF, Schulz B, Johal G, Dilkes BP (2021) Mutation of the nuclear pore complex

component, aladin1, disrupts asymmetric cell division in Zea mays (maize). G3 (Bethesda) 11(7):jkab106. https://doi.org/10.1093/g3journal/jkab106

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics 23(19):2633–2635. https://doi.org/10.1093/bioinformatics/btm308

Chang H, Hoshina N, Zhang C, Ma Y, Cao H, Wang Y, Wu DD, Bergen SE, Landén M, Hultman CM, Preisig M, Kutalik Z, Castelao E, Grigoroiu-Serbanescu M, Forstner AJ, Strohmaier J, Hecker J, Schulze TG, Müller-Myhsok B et al (2018) The protocadherin 17 gene affects cognition, personality, amygdala structure and function, synapse development and risk of major mood disorders. Mol Psychiatry 23(2):400–412. https://doi.org/10.1038/mp.2016.231

Chen E, Yu H, He J, Peng D, Zhu P, Pan S, Wu X, Wang J, Ji C, Chao Z, Xu Z, Wu Y, Chao D, Wu Y, Zhang Z (2023) The transcription factors ZmNAC128 and ZmNAC130 coordinate with Opaque2 to promote endosperm filling in maize. Plant Cell:koad215. https://doi.org/10.1093/plcell/koad215

Chen J, Wang Y, Wang F, Yang J, Gao M, Li C, Liu Y, Liu Y, Yamaji N, Ma JF, Paz-Ares J, Nussaume L, Zhang S, Yi K, Wu Z, Wu P (2015b) The rice CK2 kinase regulates trafficking of phosphate transporters in response to phosphate levels. Plant Cell 27(3):711–723. https://doi.org/10.1105/tpc.114.135335

Chen M, Zhang B, Li C, Kulaveerasingam H, Chew FT, Yu H (2015a) TRANSPARENT TESTA GLABRA1 regulates the accumulation of seed storage reserves in *Arabidopsis*. Plant Physiol 169(1):391–402. https://doi.org/10.1104/pp.15.00943

Chen W, Chen L, Zhang X, Yang N, Guo J, Wang M, Ji S, Zhao X, Yin P, Cai L, Xu J, Zhang L, Han Y, Xiao Y, Xu G, Wang Y, Wang S, Wu S, Yang F et al (2022) Convergent selection of a WD40 protein that enhances grain yield in maize and rice. Science 375(6587):eabg7985. https://doi.org/10.1126/science.abg7985

Cross FR, Umen JG (2015) The Chlamydomonas cell cycle. Plant J 82(3):370–392. https://doi.org/10.1111/tpj.12795

Dante RA, Larkins BA, Sabelli PA (2014) Cell cycle control and seed development. Front Plant Sci 23(5):493. https://doi.org/10.3389/fpls.2014.00493

Ding L, Zhao K, Zhang X, Song A, Su J, Hu Y, Zhao W, Jiang J, Chen F (2019) Comprehensive characterization of a floral mutant reveals the mechanism of hooked petal morphogenesis in Chrysanthemum morifolium. Plant Biotechnol J 17(12):2325–2340. https://doi.org/10.1111/pbi.13143

Dong Q, Xu Q, Kong J, Peng X, Zhou W, Chen L, Wu J, Xiang Y, Jiang H, Cheng B. (2019) Overexpression of ZmbZIP22 gene alters endosperm starch content and composition in maize and rice. 283:407-415. https://doi.org/10.1016/j.plantsci.2019.03.001.

Dossa K, Zhou R, Li D, Liu A, Qin L, Mmadi MA, Su R, Zhang Y, Wang J, Gao Y, Zhang X, You J (2021) A novel motif in the 5'-UTR of an orphan gene 'Big Root Biomass' modulates root biomass in sesame. Plant Biotechnol J 19(5):1065–1079. https://doi.org/10.1111/pbi.13531

Everett LJ, Huang W, Zhou S, Carbone MA, Lyman RF, Arya GH, Geisz MS, Ma J, Morgante F, St Armour G, Turlapati L, Anholt RRH, Mackay TFC (2020) Gene expression networks in the *Drosophila* Genetic Reference Panel. Genome Res 30(3):485–496. https://doi.org/10.1101/gr.257592.119

Finegan C, Boehlein SK, Leach KA, Madrid G, Hannah LC, Koch KE, Tracy WF, Resende MFR Jr (2022) Genetic perturbation of the starch biosynthesis in maize endosperm reveals sugar-responsive gene networks. Front Plant Sci 12:800326. https://doi.org/10.3389/fpls.2021.800326

Fu FF, Xue HW (2010) Co-expression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. Plant Physiol 154(2):927–938. https://doi.org/10.1104/pp.110.159517

Gong J, Schumacher F, Lim U, Hindorff LA, Haessler J, Buyske S, Carlson CS, Rosse S, Bůžková P, Fornage M, Gross M, Pankratz N, Pankow JS, Schreiner PJ, Cooper R, Ehret G, Gu CC, Houston D, Irvin MR et al (2013) Fine mapping and identification of BMI loci in African Americans. Am J Hum Genet 93(4):661–671. https://doi.org/10.1016/j.ajhg.2013.08.012

Guo L, Jiang L, Lu XL, Liu CM (2018) ANAPHASE PROMOTIN COMPLEX/CYCLOSOME-mediated cyclin B1 degradation is critical for cell cycle synchronization in syncytial endosperms. J Integr Plant Biol 60(6):448–454. https://doi.org/10.1111/jipb.12641

Guo Z, Yang Q, Huang F, Zheng H, Sang Z, Xu Y, Zhang C, Wu K, Tao J, Prasanna BM, Olsen MS, Wang Y, Zhang J, Xu Y (2021) Development of high-resolution multiple-SNP arrays for genetic analyses and molecular breeding through genotyping by target sequencing and liquid chip. Plant Commun 2(6):100230. https://doi.org/10.1016/j.xplc.2021.100230

Han X, Zhang D, Hao H, Luo Y, Zhu Z, Kuai B (2023) Transcriptomic analysis of three differentially senescing maize (Zea mays L.) inbred lines upon heat stress. Int J Mol Sci 24(12):9782. https://doi.org/10.3390/ijms24129782

Hoopes GM, Hamilton JP, Wood JC, Esteban E, Pasha A, Vaillancourt B, Provart NJ, Buell CR (2019) An updated gene atlas for maize reveals organ-specific and stress-induced genes. Plant J 97(6):1154–1167. https://doi.org/10.1111/tpj.14184

Hu S, Wang M, Zhang X, Chen W, Song X, Fu X, Fang H, Xu J, Xiao Y, Li Y, Bai G, Li J, Yang X (2021) Genetic basis of kernel starch content decoded in a maize multi-parent population. Plant Biotechnol J 19(11):2192–2205. https://doi.org/10.1111/pbi.13645

Jeon JS, Ryoo N, Hahn TR, Walia H, Nakamura Y (2010) Starch biosynthesis in cereal endosperm. Plant Physiol Biochem 48(6):383–392. https://doi.org/10.1016/j.plaphy.2010.03.006

Kaushik RP, Khush GS (1991) Genetic analysis of endosperm mutants in rice Oryza sativa L. Theor Appl Genet 83(2):146–152. https://doi.org/10.1007/BF00226243

Kumar S, Li G, Huang X, Ji Q, Zhou K, Hou H, Ke W, Yang J (2021) Phenotypic, nutritional, and antioxidant characterization of blanched *Oenanthe javanica* for preferable cultivar. Front Plant Sci 12:639639. https://doi.org/10.3389/fpls.2021.639639

Leiva-Neto JT, Grafi G, Sabelli PA, Dante RA, Woo YM, Maddock S, Gordon-Kamm WJ, Larkins BA (2004) A dominant negative mutant of cyclin-dependent kinase A reduces endoreduplication but not cell size or gene expression in maize endosperm. Plant Cell 16(7):1854–1869. https://doi.org/10.1105/tpc.022178

Li C, Huang Y, Huang R, Wu Y, Wang W (2018b) The genetic architecture of amylose biosynthesis in maize kernel. Plant Biotechnol J 16(2):688–695. https://doi.org/10.1111/pbi.12821

Li C, Yue Y, Chen H, Qi W, Song R (2018a) The ZmbZIP22 transcription factor regulates 27-kD γ-zein gene transcription during maize endosperm development. Plant Cell 30(10):2402–2424. https://doi.org/10.1105/tpc.18.00422

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N, Liu J, Warburton ML, Cheng Y, Hao X, Zhang P, Zhao J, Liu Y, Wang G, Li J, Yan J (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet 45(1):43–50. https://doi.org/10.1038/ng.2484

Li M, Zhang YW, Xiang Y, Liu MH, Zhang YM (2022) IIIVmrMLM: the R and C++ tools associated with 3VmrMLM, a comprehensive GWAS method for dissecting quantitative traits. Mol Plant 15(8):1251–1253. https://doi.org/10.1016/j.molp.2022.06.002

Liu G, Wu Y, Xu M, Gao T, Wang P, Wang L, Guo T, Kang G (2016b) Virus-induced gene silencing identifies an important role of the TaRSR1 transcription factor in starch synthesis in bread wheat. Int J Mol Sci 17(10):1557. https://doi.org/10.3390/ijms17101557

Liu J, Wu MW, Liu CM (2022) Cereal endosperms: development and storage product accumulation. Annu Rev Plant Biol 73:255–291. https://doi.org/10.1146/annurev-arplant-070221-024405

Liu N, Xue Y, Guo Z, Li W, Tang J (2016a) Genome-wide association study identifies candidate genes for starch content regulation in maize kernels. Front Plant Sci 7:1046. https://doi.org/10.3389/fpls.2016.01046

McGeachie MJ, Wu AC, Tse SM, Clemmer GL, Sordillo J, Himes BE, Lasky-Su J, Chase RP, Martinez FD, Weeke P, Shaffer CM, Xu H, Denny JC, Roden DM, Panettieri RA Jr, Raby BA, Weiss ST, Tantisira KG (2015) CTNNA3 and SEMA3D: promising loci for asthma exacerbation identified through multiple genome-wide association studies. J Allergy Clin Immunol 136(6):1503–1510. https://doi.org/10.1016/j.jaci.2015.04.039

Neff MM, Turk E, Kalishman M (2002) Web-based primer design for single nucleotide polymorphism analysis. Trends Genet 18(12):613–615. https://doi.org/10.1016/s0168-9525(02)02820-2

Pal S, Gupta R, Kim H, Wickramasinghe P, Baubet V, Showe LC, Dahmane N, Davuluri RV (2011) Alternative transcription exceeds alternative splicing in generating the transcriptome diversity of cerebellar development. Genome Res 21(8):1260–1272. https://doi.org/10.1101/gr.120535.111

Qin P, Zhang G, Hu B, Wu J, Chen W, Ren Z, Liu Y, Xie J, Yuan H, Tu B, Ma B, Wang Y, Ye L, Li L, Xiang C, Li S (2021) Leaf-derived ABA regulates rice seed development via a transporter-mediated and temperature-sensitive mechanism. Sci Adv 7(3):eabc8873. https://doi.org/10.1126/sciadv.abc8873

Sabelli PA, Liu Y, Dante RA, Lizarraga LE, Nguyen HN, Brown SW, Klingler JP, Yu J, LaBrant E, Layton TM, Feldman M, Larkins BA (2013) Control of cell proliferation, endoreduplication, cell size, and cell death by the retinoblastoma-related pathway in maize endosperm. Proc Natl Acad Sci USA 110(19):E1827–E1836. https://doi.org/10.1073/pnas.1304903110

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Setter TL, Yan J, Warburton M, Ribaut JM, Xu Y, Sawkins M, Buckler ES, Zhang Z, Gore MA (2011) Genetic association mapping identifies single nucleotide polymorphisms in genes that affect abscisic acid levels in maize floral tissues during drought. J Exp Bot 62(2):701–716. https://doi.org/10.1093/jxb/erq308

She KC, Kusano H, Koizumi K, Yamakawa H, Hakata M, Imamura T, Fukuda M, Naito N, Tsurumaki Y, Yaeshima M, Tsuge T, Matsumoto K, Kudoh M, Itoh E, Kikuchi S, Kishimoto N, Yazaki J, Ando T, Yano M et al (2010) A novel factor FLOURY ENDOSPERM2 is involved in regulation of rice grain size and starch quality. Plant Cell 22(10):3280–3294. https://doi.org/10.1105/tpc.109.070821

Song L, Yu D, Zheng H, Wu G, Sun Y, Li P, Wang J, Wang C, Lv B, Tang X (2021) Weighted gene co-expression network analysis unveils gene networks regulating folate biosynthesis in maize endosperm. 3 Biotech 11(10):441. https://doi.org/10.1007/s13205-021-02974-7

Song Y, Luo G, Shen L, Yu K, Yang W, Li X, Sun J, Zhan K, Cui D, Liu D, Zhang A (2020) TubZIP28, a novel bZIP family transcription factor from Triticum urartu, and TabZIP28, its homologue from Triticum aestivum, enhance starch synthesis in wheat. New Phytol 226(5):1384–1398. https://doi.org/10.1111/nph.16435

Sreenivasulu N, Radchuk V, Alawady A, Borisjuk L, Weier D, Staroske N, Fuchs J, Miersch O, Strickert M, Usadel B, Wobus U, Grimm B, Weber H, Weschke W (2010) De-regulation of abscisic acid contents causes abnormal endosperm development in the barley mutant seg8. Plant

J 64(4):589–603. https://doi.org/10.1111/j.1365-313X.2010.04350.x

Sun G, Zhang X, Duan H, Gao J, Li N, Su P, Xie H, Li W, Fu Z, Huang Y, Tang J (2022) Dissection of the genetic architecture of peduncle vascular bundle-related traits in maize by a genome-wide association study. Plant Biotechnol J 20(6):1042–1053. https://doi.org/10.1111/pbi.13782

Tedja MS, Wojciechowski R, Hysi PG, Eriksson N, Furlotte NA, Verhoeven VJM, Iglesias AI, Meester-Smoor MA, Tompson SW, Fan Q, Khawaja AP, Cheng CY, Höhn R, Yamashiro K, Wenocur A, Grazal C, Haller T, Metspalu A, Wedenoja J et al (2018) Genome-wide association meta-analysis highlights light-induced signaling as a driver for refractive error. Nat Genet 50(6):834–848. https://doi.org/10.1038/s41588-018-0127-7

Tetlow IJ (2011) Starch biosynthesis in developing seeds. Seed Sci Res 21(1):5–32

Wang T, Wang M, Hu S, Xiao Y, Tong H, Pan Q, Xue J, Yan J, Li J, Yang X (2015) Genetic basis of maize kernel starch content revealed by high-density single nucleotide polymorphism markers in a recombinant inbred line population. BMC Plant Biol 15:288. https://doi.org/10.1186/s12870-015-0675-2

Wang Y, Liu W, Wang H, Du Q, Fu Z, Li WX, Tang J (2020) ZmEHD1 is required for kernel development and vegetative growth through regulating auxin homeostasis. Plant Physiol 182(3):1467–1480. https://doi.org/10.1104/pp.19.01336

Wu B, Chang H, Marini R, Chopra S, Reddivari L (2021) Characterization of maize near-isogenic lines with enhanced flavonoid expression to be used as tools in diet-health complexity. Front Plant Sci 11:619598. https://doi.org/10.3389/fpls.2020.619598

Wu J, Chen L, Chen M, Zhou W, Dong Q, Jiang H, Cheng B (2019) The DOF-domain transcription factor ZmDOF36 positively regulates starch synthesis in transgenic maize. Front Plant Sci 10:465. https://doi.org/10.3389/fpls.2019.00465

Wu MW, Zhao H, Zhang JD, Guo L, Liu CM (2020) RADICLELESS 1 (RL1)-mediated nad4 intron 1 splicing is crucial for embryo and endosperm development in rice (Oryza sativa L.). Biochem Biophys Res Commun 523(1):220–225. https://doi.org/10.1016/j.bbrc.2019.11.084

Wu X, Liu J, Li D, Liu CM (2016) Rice caryopsis development II: dynamic changes in the endosperm. J Integr Plant Biol 58(9):786–798. https://doi.org/10.1111/jipb.12488

Xiao Q, Wang Y, Du J, Li H, Wei B, Wang Y, Li Y, Yu G, Liu H, Zhang J, Liu Y, Hu Y, Huang Y (2017) ZmMYB14 is an important transcription factor involved in the regulation of the activity of the ZmBT1 promoter in starch biosynthesis in maize. FEBS J 284(18):3079–3099. https://doi.org/10.1111/febs.14179

Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, Kubo M, Nakamura Y, Kamatani N (2008) Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. Am J Hum Genet 83(4):445–456. https://doi.org/10.1016/j.ajhg.2008.08.019

Yang N, Lu Y, Yang X, Huang J, Zhou Y, Ali F, Wen W, Liu J, Li J, Yan J (2014) Genome wide association studies using a new nonparametric model reveal the genetic architecture of 17 agronomic traits in an enlarged maize association panel. PLoS Genet 10(9):e1004573. https://doi.org/10.1371/journal.pgen.1004573

Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Li K, Yang N, Li Y, Zhong T, Li J, Chen Y, Yan J, Yang X, Xu M (2013) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc Natl Acad Sci USA 110(42):16969–16974. https://doi.org/10.1073/pnas.1310949110

Yang X, Gao S, Xu S et al (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. Mol Breed 28(4):511–526

Yang X, Yan J, Shah T, Warburton ML, Li Q, Li L, Gao Y, Chai Y, Fu Z, Zhou Y, Xu S, Bai G, Meng Y, Zheng Y, Li J (2010) Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. Theor Appl Genet 121(3):417–431. https://doi.org/10.1007/s00122-010-1320-y

Yin W, Xiao Y, Niu M, Meng W, Li L, Zhang X, Liu D, Zhang G, Qian Y, Sun Z, Huang R, Wang S, Liu CM, Chu C, Tong H (2020) ARGONAUTE2 enhances grain length and salt tolerance by activating *BIG GRAIN3* to modulate cytokinin distribution in rice. Plant Cell 32(7):2292–2306. https://doi.org/10.1105/tpc.19.00542

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38(2):203–208. https://doi.org/10.1038/ng1702

Zhang C, Zhu J, Chen S, Fan X, Li Q, Lu Y, Wang M, Yu H, Yi C, Tang S, Gu M, Liu Q (2019b) Wx^{lv}, the ancestral allele of rice waxy gene. Mol Plant 12(8):1157–1166. https://doi.org/10.1016/j.molp.2019.05.011

Zhang F, Wu J, Sade N, Wu S, Egbaria A, Fernie AR, Yan J, Qin F, Chen W, Brotman Y, Dai M (2021) Genomic basis underlying the metabolome-mediated drought adaptation of maize. Genome Biol 22(1):260. https://doi.org/10.1186/s13059-021-02481-1

Zhang H, Lang Z, Zhu JK (2018) Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol 19(8):489–506. https://doi.org/10.1038/s41580-018-0016-z

Zhang X, Wang M, Zhang C, Dai C, Guan H, Zhang R (2022) Genetic dissection of QTLs for starch content in four maize DH populations. Front Plant Sci 13:950664. https://doi.org/10.3389/fpls.2022.950664

Zhang X, Warburton ML, Setter T, Liu H, Xue Y, Yang N, Yan J, Xiao Y (2016b) Genome-wide association studies of drought-related metabolic changes in maize using

an enlarged SNP panel. Theor Appl Genet 129(8):1449–1463. https://doi.org/10.1007/s00122-016-2716-0

Zhang XF, Tong JH, Bai AN, Liu CM, Xiao LT, Xue HW (2020) Phytohormone dynamics in developing endosperm influence rice grain shape and quality. J Integr Plant Biol 62(10):1625–1637. https://doi.org/10.1111/jipb.12927

Zhang Z, Dong J, Ji C, Wu Y, Messing J (2019a) NAC-type transcription factors regulate accumulation of starch and protein in maize seeds. Proc Natl Acad Sci USA 116(23):11223–11228. https://doi.org/10.1073/pnas.1904995116

Zhang Z, 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(4):355–360. https://doi.org/10.1038/ng.546

Zhang Z, Zheng X, Yang J, Messing J, Wu Y (2016a) Maize endosperm-specific transcription factors O2 and PBF network the regulation of protein and starch synthesis. Proc Natl Acad Sci USA 113(39):10842–10847. https://doi.org/10.1073/pnas.1613721113

Zhao Y, Zhou DX (2012) Epigenomic modification and epigenetic regulation in rice. J Genet Genomics 39(7):307–315. https://doi.org/10.1016/j.jgg.2012.02.009

Zheng XM, Chen J, Pang HB, Liu S, Gao Q, Wang JR, Qiao WH, Wang H, Liu J, Olsen KM, Yang QW (2019) Genome-wide analyses reveal the role of noncoding variation in complex traits during rice domestication. Sci Adv 5(12):eaax3619. https://doi.org/10.1126/sciadv.aax3619

Zhou YH, Li G, Zhang YM (2022) A compressed variance component mixed model framework for detecting small and linked QTL-by-environment interactions. Brief Bioinform 23(2):bbab596. https://doi.org/10.1093/bib/bbab596

Zhu Y, Cai XL, Wang ZY, Hong MM (2003) An interaction between a MYC protein and an EREBP protein is involved in transcriptional regulation of the rice Wx gene. J Biol Chem 278(48):47803–47811. https://doi.org/10.1074/jbc.M302806200