

## Article

# The Identification of a Yield-Related Gene Controlling Multiple Traits Using GWAS in Sorghum (*Sorghum bicolor* L.)

Yizhong Zhang<sup>1,2,3</sup>, Xinqi Fan<sup>2,3</sup>, Du Liang<sup>2,3</sup>, Qi Guo<sup>2,3</sup>, Xiaojuan Zhang<sup>2,3</sup>, Mengen Nie<sup>2,3</sup>, Chunhong Li<sup>4</sup>, Shan Meng<sup>4</sup>, Xiangui Zhang<sup>4</sup>, Peng Xu<sup>4</sup>, Wenqi Guo<sup>4</sup>, Huiyan Wang<sup>1,2,3</sup>, Qingshan Liu<sup>3,\*</sup> and Yuxiang Wu<sup>1,\*</sup>

<sup>1</sup> College of Agriculture, Shanxi Agricultural University, Taigu 030801, China

<sup>2</sup> Sorghum Research Institute, Shanxi Key Laboratory of Sorghum Genetic and Germplasm Innovation, Shanxi Agricultural University, Yuci 030600, China

<sup>3</sup> Shanxi Key Laboratory of Minor Crops Germplasm Innovation and Molecular Breeding, State Key Laboratory of Sustainable Dryland Agriculture (In Preparation), Shanxi Agricultural University, Taiyuan 030031, China

<sup>4</sup> Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China

\* Correspondence: 673673@126.com (Q.L.); yuxiangwu2009@hotmail.com (Y.W.)

**Abstract:** *Sorghum bicolor* (L.) is one of the oldest crops cultivated by human beings which has been used in food and wine making. To understand the genetic diversity of sorghum breeding resources and further guide molecular-marker-assisted breeding, six yield-related traits were analyzed for 214 sorghum germplasm from all over the world, and 2,811,016 single-nucleotide polymorphisms (SNPs) markers were produced by resequencing these germplasms. After controlling Q and K, QTLs were found to be related to the traits using three algorithms. Interestingly, an important QTL was found which may affect multiple traits in this study. It was the most likely candidate gene for the gene *SORBI\_3008G116500*, which was a homolog of *Arabidopsis thaliana* gene-*VIP5* found by analyzing the annotation of the gene in the LD block. The haplotype analysis showed that the *SORBI\_3008G116500*<sup>hap3</sup> was the elite haplotype, and it only existed in Chinese germplasms. The traits were proven to be more associated with the SNPs of the *SORBI\_3008G116500* promoter through gene association studies. Overall, the QTLs and the genes identified in this study would benefit molecular-assisted yield breeding in sorghum.



**Citation:** Zhang, Y.; Fan, X.; Liang, D.; Guo, Q.; Zhang, X.; Nie, M.; Li, C.; Meng, S.; Zhang, X.; Xu, P.; et al. The Identification of a Yield-Related Gene Controlling Multiple Traits Using GWAS in Sorghum (*Sorghum bicolor* L.). *Plants* **2023**, *12*, 1557. <https://doi.org/10.3390/plants12071557>

Academic Editor: Othmane Merah

Received: 24 February 2023

Revised: 29 March 2023

Accepted: 30 March 2023

Published: 4 April 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** sorghum; six yield-related traits; GWAS; *VIP5*; breeding

## 1. Introduction

*Sorghum bicolor* is one of the oldest crops cultivated by human beings, and it is also one of the only crops affected by artificial selection evolution [1]. Meanwhile, the wild sorghum that is commonly distributed in Africa is a promising resource for sorghum improvement [2]. *Sorghum* is widely distributed all over the world and has various uses, such as in food, wine making, and so on [3]. Moreover, sweet sorghum could be used for bioenergy with genetically diverse characteristics and variation exits [4]. Sweet sorghum was considered in the model system for bioenergy crops due to its favorable low-input cost traits [5]. Due to the different climate, soil setting, and cultivation system in the cultivation area, various types in sorghum have formed different agronomic, physiological, and biochemical characteristics through natural and artificial selection in the long-term cultivation process [6]. The variations of traits laid a foundation for the exploration of related genes in sorghum.

With the increase in the world population, the demand for grain production is also higher [7]. Understanding the genetic variations in traits related to yield was a necessary basis for breeding [8]. In the process of breeding, yield has always been the concern of breeders, so analyzing the traits relating to yield could provide effective guidance to breeders [9].

To date, the traits related to yield have been well studied in plants. In rice, plant height and the number of internodes are the key factors determining not only plant biomass, but also yield. For instance, the *GRF (GROWTH-REGULATED FACTOR)4* gene increased plant height and yield [10]; the panicle type also directly affects yield. The *DEP (DENSE AND ERECT PANICLE)1* gene of rice, which encodes a phosphatidylethanolamine-binding protein-like domain protein, could directly improve yield by affecting the panicle type [11]. Meanwhile, the heading stage can also affect the yield in rice. The *DTH (Days to heading)8* gene was isolated through the heading stage. At the same time, the *DTH8* superior haplotype could improve the yield of rice [12]. Moreover, the 100-grain weight is one of the most direct factors in yield. Many genes in plants such as *TGW (THOUSAND-GRAIN WEIGHT)6* could increase the yield [13]. Previous studies showed that the traits related to yield could be identified, and that the traits would be better for explaining the mechanism of yield. In sorghum, some genes which influenced the yield were cloned. The *BY (BIOMASS YIELD)1* gene encodes a 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (DAHPS), which was cloned by map-based cloning and could regulate biomass and grain yield through the shikimate pathway [14]. Allelic variations in the NAC transcription factor were found to affect grain yield, sugar yield, and biomass composition in near-isogenic lines (NILs) [15]. The previous study showed that other traits, such as the leaf length, leaf breadth, number of leaves per plant, panicle weight, and hundred-seed weight may have also been given importance along with yield according to correlation analysis results [16]. Therefore, yield was a complex trait, and understanding it required inheritance variations in multiple components to be analyzed together.

Genome-wide association study (GWAS) aims to explore excellent allelic variation using natural variation populations and provide superior natural variations in materials for breeders [17–20]. Compared to bi-parental quantitative trait loci (QTL) analysis, GWAS could detect multiple alleles, save time, and offer some other advantages [8], such as the deconstruction of offspring populations. One natural variation population could be used to investigate multiple traits, and it effectively solved the problem that multiple components of yield have different genetic variations [21–23].

Meanwhile, GWASs have been widely used in sorghum. In order to improve environmental adaptation, 1943 georeferenced sorghum landraces were used for the SNP detection of experimental drought stress and aluminum toxicity [24]. Recently, researchers observed major heat tolerance traits and performed GWASs for traits responsive to heat stress using 374 accessions [25]. Association mapping was used for germinability and seedling vigor under low-temperature conditions in sorghum. Besides the traits related to abiotic stress, the traits with seed were studied using association mapping. Li et al. used the genotyping by sequencing (GBS) method to genotype 238 mini-core collection sorghum and 7 breeding varieties, and they collected the data of forage quality-related traits. As a result, 42 SNPs were found which were associated with traits and 14 genes were predicted [26]. The grain polyphenol concentrations were studied using the global germplasm, and it was found that the flavonoid pathway loci may be useful to guide the future enhancement of cereal polyphenols [27]. To identify the natural variations in sorghum grain quality traits, 196 diverse sorghum inbred lines were used for GWASs, and 1,414,494 genetic loci were found [28]. Three traits (grain yield per primary panicle, grain number per primary panicle, and 1000-grain weight) related to yield were collected from 390 diverse accessions and were used for GWASs. As a result, the gene, namely the ethylene receptor, was predicted as the candidate gene [29]. The other trait, the nodal root angle, was analyzed using GWAS in sorghum and 58.2% of the phenotypic variance was explained [30]. In sweet sorghum, important traits, including brix and height, were conducted for association mapping [31]. The previous study above showed that GWAS could be useful to analyze the population and detect the significant SNPs in sorghum.

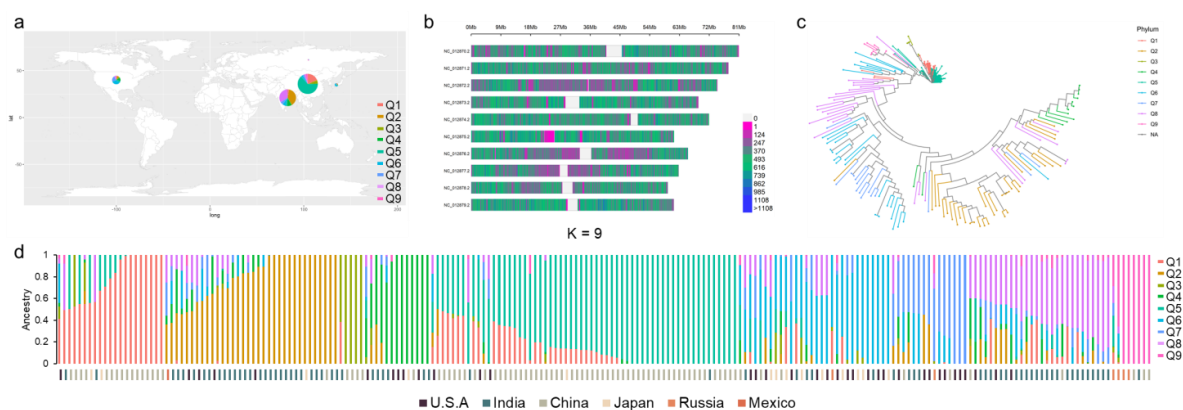
Although there are some studies on the traits of yield, the genetic variation in other specific yield traits, such as plant height and the relationship between genetic variations with various traits, remains largely unknown. We herein collected a population consisting

of 214 landraces from diverse countries to observe six traits related to yield. Via the resequencing method, 2,811,016 markers were called. Using the emmax, fast-lmm, and lmm algorithms with controlling Q and K, we identified numbers of SNP markers associated with the traits. Finally, the *SORBI\_3008G116500* gene was found, which may control four traits through the annotation of genes in the LD block, and the elite haplotype was analyzed.

## 2. Result

### 2.1. Population Collection and Genotyping

The landraces from different regions could provide more genetic variation; therefore, 214 landraces from China (86), India (71), America (36), Japan (14), Russia (6), and Mexico (1) were collected (Figure 1a, Table S1). The landraces laid the foundation of genetic variation for the GWAS. After that, the population was genotyped using the resequencing method. As a result, we obtained about 2198 G raw base data, and the Q30 percentage was 92.5%. The raw reads were 14.6 Gb. The average sequencing depth was 13.4, and the average coverage ratio of 10 times of population was 71.2%. After analysis, we obtained 6,738,606 SNPs and 1,336,590 indels. The transition–transversion rate was about 1.9 in all varieties, and the average heterozygosity was 21.4%. Overall, the high-quality molecular markers derived from the germplasm were obtained based on resequencing and analyzing data, which then laid the foundation for subsequent analysis.



**Figure 1.** The analysis of population. (a) The distribution of all 214 landraces. (b) The density of SNPs in all chromosomes. (c) The neighbor–joining tree of all landraces. The colored subsections within each vertical bar indicate membership coefficient (Q) of the accession to different clusters. (d) The populations were pre–determined with  $K = 9$  based on ADMIXTURE analysis. The bottom was the country of origin for each landrace. Different color indicates different membership coefficient (Q) in (a,c,d).

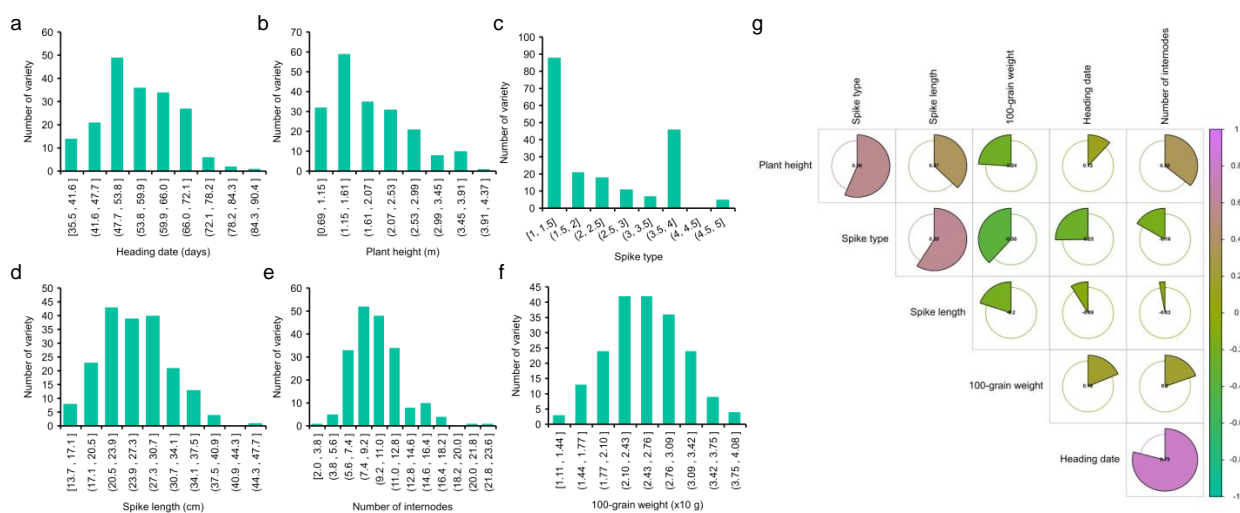
### 2.2. SNP Markers Selection and Population Structure Analysis

Firstly, 2,811,016 SNP markers were left after filtering with the missing rate ( $<0.8$ ) and the minor allele frequency ( $>0.05$ ). Through separating the markers into each chromosome, we found that the average number of SNP markers was about 281,102. The Chr.NC\_012871 had the most SNPs (327,237), while chr.NC\_012876 had the least SNPs (224,286) (Figure 1b, Table S2). Based on the SNP markers, including the population structure, phylogenetic tree, PCA, and kinship analysis, admixture software was used, and the cross-validation (CV) error was calculated (Figure 1d, Figure S1). The results show that the CV error was the smallest when  $K = 9$ . The population could therefore be divided into nine subpopulations (Figure 1d). Combined with the origin of the landraces, the landrace from China showed the highest proportion in Q1, Q3, and Q5, while the landraces from India were mainly found in Q2 and Q8 (Figure 1d). Moreover, the neighbor-joining clustering of landraces based on genetic distance was analyzed, and the result was almost consistent with that of population structure (Figure 1c). Then, most varieties had distant phylogenetic relationships according to the kinship analysis (Figure S2). Lastly, PCA analysis showed almost no varieties with

the same background (Figure S3, Table S3). The above results show that the population we selected was suitable for GWASs and could be used to identify the number of variants in some loci.

### 2.3. Phenotype Analysis

In order to elucidate the genetic base of traits related to the yield of sorghums, traits including the heading date, plant height, spike type (Figure S4), spike length, number of internodes, and 100-grain weight were observed (Figure 2a–f). As a result, extensive phenotypic variations were observed in all traits. The variation coefficients in the six traits were 0.21, 0.40, 0.53, 0.28, 0.30, and 0.22. The phenotype values of plant height were 0.69 to 3.96. The spike type ranged from 1 to 5, while the spike length varied from 13.67 to 69.17. Lastly, the number of sections ranged from 2 to 22 (Table S4, Table 1). After calculating the correlation of each trait, the number of internodes had the most relevance with the heading date, while the smallest relevance was found with the spike length (Figure 2g). The results show that the total observed six traits were abundant, but the intrinsic mechanism relating to yield formation may be different.



**Figure 2.** The analysis of phenotype. (a–f) The frequency distribution histogram of six traits. (g) The relevance of each trait.

**Table 1.** The minimum (Min), maximum (Max), mean values (Mean), and coefficient of variation (CV) of heading date (days), plant height (m), spike type, spike length (cm), number of internodes, and 100-grain weight (g) in the population.

Trait	Min	Max	Mean	CV
Heading date (days)	35.50	104.90	57.69	0.21
Plant height (m)	0.69	3.96	1.90	0.40
Spike type	1.00	5.00	2.34	0.53
Spike length (cm)	13.67	69.17	26.75	0.28
Number of internodes	2.00	22.00	9.79	0.30
100-grain weight ( $\times 10$ g)	1.11	4.03	2.56	0.22

### 2.4. LD Decay Analysis and GWAS

To estimate the linkage disequilibrium (LD) of the population, the  $r^2$  was calculated using the Plink software. The variation in LD decay ranged from 9.9 to 70.2 kb, and the average was 17.4 K. The greatest LD decay in all chromosomes was found in NC\_012875, while the least was found in NC\_012870 (Figure S5).

Based the collection of phenotype data and the population analysis, three GWAS algorithms were used to detect the QTLs. As results, 29 SNPs which exceeded the

threshold ( $-\log_{10}(p) = 7.75$ ) were observed in all six traits using the emmax (Figure 3, Table S5). The greatest number of SNPs (21) was detected in spike length trait, while the smallest numbers of SNPs (0) were found in plant height, the number of internodes, and 100-grain weight trait. Of the 21 SNPs significantly associated with spike length, 3 were found on chromosome NC\_012871.2, 1 was found on chromosome NC\_012872.2, 7 were found on chromosome NC\_012874.2, 6 were found on chromosome NC\_012876.2, 1 was found on chromosome NC\_012878.2, and 3 were found on chromosome NC\_012879.2. The other three and five significant SNPs on chromosome NC\_012875.2 were found in the heading date trait and on chromosome NC\_012874.2 in the spike-type trait, respectively (Table S5). With the calculation of the fastlmm model, in total, 1175 SNPs were found (Figure S6, Table S6). The same positions were detected in the heading date and the spike-type trait. Among all the significant SNPs, the NC\_012874.2 chromosome had the highest spike length (1079) (Table S6). The SNP (NC\_012874.2 1992075) was found in the plant height trait, while no SNP was found using emmax. Moreover, the lmm model of GEMMA was used to identify the 1320 associated SNPs (Figure S7, Table S7). Luckily, there were two significant SNPs in chromosomes NC\_012876.2 and NC\_012878.2 (Table S7), while no SNPs were found in the other two algorithms (Tables S5 and S6).



**Figure 3.** The Manhattan and Q-Q plot of six traits, including heading date (HD), plant height (PL), spike type (ST), spike length (SL), number of internodes (NI), and 100-grain weight (100 GW), based on the emmax algorithm. The light red dash line indicates the QTL that was identified in four traits.

The change in heading date may lead to increases in the yield [32]. By using GWAS to detect the loci controlling the heading stage, three SNPs exceeded the threshold ( $-\log_{10}(p) = 7.75$ ), representing the same QTL (Figure 3). The other loci did not exceed the threshold, and it may contain genes controlling heading date.

In the plant height trait, there was one common significant SNP (NC\_012874.2 1992075) within one QTL using the fast-lmm and lmm algorithms. From the Manhattan plot, this peak was more continuous than other peaks (Figure 3). Therefore, a gene which controls the plant height would need to be cloned in future studies.

The shape and length of spike were also directly related to the yield of plants. In the GWAS results, 5 and 1307 SNPs were found to be significantly associated with spike traits using the lmm model. In panicle length traits, multiple SNPs on the NC\_012874.2 chromosome were found to exceed the threshold (Figure 3). This indicates that there may be multiple genes related to ear length on this chromosome.

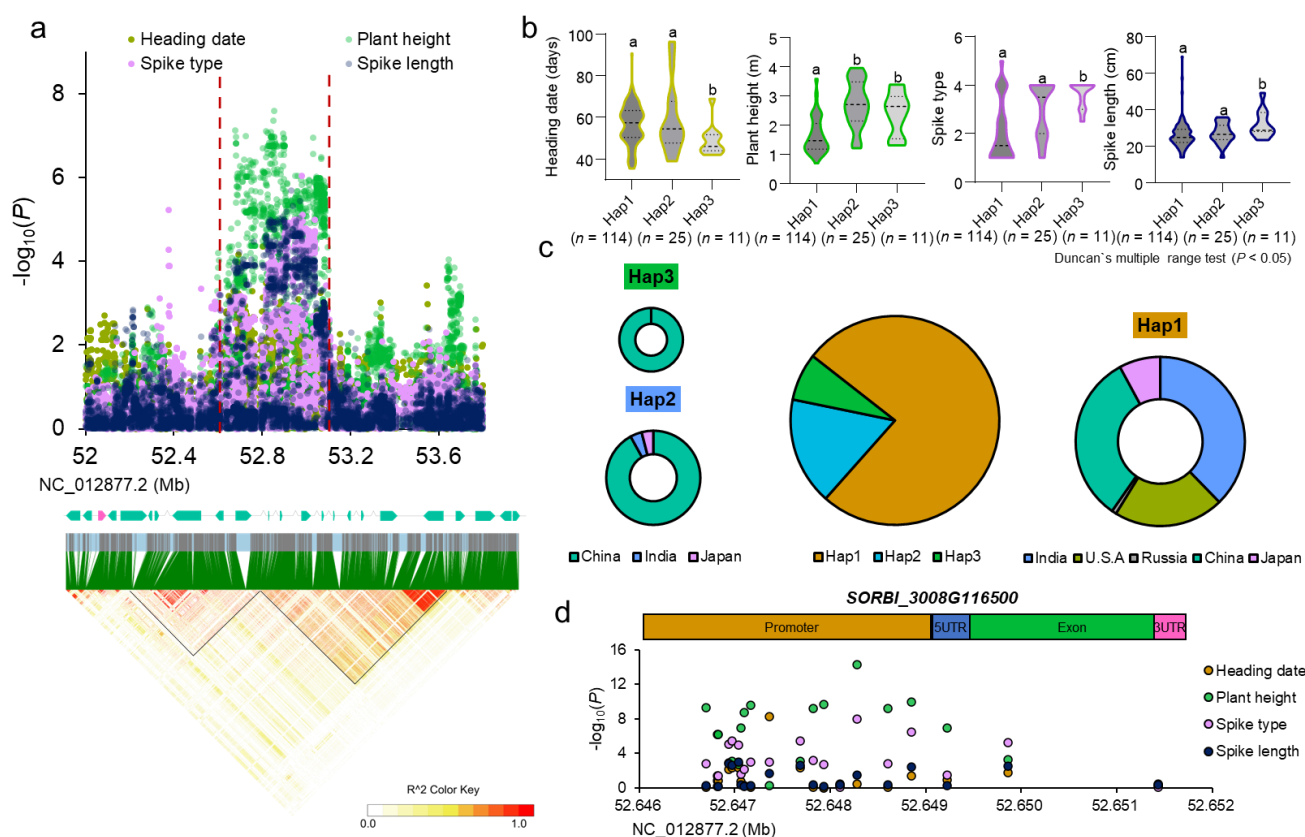
After detecting the significant SNPs in the two traits, no SNPs associated with the number of internodes and 100-grain weight were identified unfortunately. From the Q-Q plot, it is possible that the model did not match the two phenotypes (Figure S3). Although there were no significant SNPs, there were some peaks which showed the trend of further research, such as the QTL located in the end of NC\_012878 (Figure 3).

### 2.5. Prediction and Analysis of Yield-Related Gene

Although there were numbers of QTLs in the GWAS, it was important to find the QTLs which were identified in several traits. Additionally, we could clone genes that affected multiple phenotypes and explored excellent alleles. Luckily, there were signals at the adjacent position on the NC\_012877 chromosome in the traits of heading date, plant height, spike type, and spike length (Figure 4a up). There was probably a gene with multiple functions at the locus, such as the *DTH8* gene, in rice [12].

To identify whether some genes were the causal genes in the QTL, we calculated the LD block. The significant SNP was in the LD block from 52,637,266 to 52,792,992 and from 52,793,087 to 53,095,551 (Figure 4a bottom). In the LD blocks, there were 23 predicted genes (Figure 4a middle, Table S8). In all genes, the *SORBI\_3008G116500* was annotated as *VIP5*. Luckily, the homologue gene of *SORBI\_3008G116500* in *Arabidopsis thaliana* was reported to affect the heading date [33,34]. So, the *SORBI\_3008G116500* gene was mostly the candidate gene.

To observe the elite haplotype, three haplotypes were found according to the SNP in the promoter and gene (Table S9). The haplotype analysis showed that hap3 was superior compared to the other two based on the four traits' data (Figure 4b). In all varieties, Hap3 provided the least percentages compared with the other two haplotypes, suggesting that the elite haplotype was not fully used (Figure 4c). Combined with the origin of the three haplotypes, the Hap1 almost spread all over the regions where we collected the germplasm. The Hap2 was most present in China, while the elite haplotype-Hap3 only existed in China (Figure 4c). This indicates that higher polymorphism in the candidate gene was found in China. Through the gene association study, the SNP in the promoter may play a more significant role in the gene function (Figure 4d). Above that, we believed that *SORBI\_3008G116500* was most likely the candidate gene in the QTL and the superior haplotype could be widely used for sorghum breeding, especially in other countries.



**Figure 4.** The analysis of the QTL. (a) Local Manhattan plot (top), the gene in the LD block (middle), and LD heatmap (bottom) surrounding the peak on chromosome NC\_012877.2. The red dashed line indicates the candidate region for the peak. (b) The haplotype analysis of the *SORBI\_3008G116500* gene. Different letters represent significant differences ( $P < 0.05$ ) using Duncan's multiple range test. (c) The geographical distribution of different haplotypes. (d) The gene association study of *SORBI\_3008G116500*.

### 3. Discussion

#### 3.1. GWAS Results and Observed Traits

In our study, the most significant SNPs were detected in the spike length trait using three algorithms. Meanwhile, the panicle length has been investigated by linkage mapping [35]. Compared with the previous QTLs, more novel QTLs were found in chromosome NC\_012874.2 and so on (Tables S5–S7). At the same time, GWAS was also carried out on the grains per panicle and grain size in sorghum [36]. In our study, many QTLs were detected using GWAS, which improved the knowledge on the genetic variations in yield, but more QTLs would be found if we observed more traits related to yield such as grain size [37] or traits related to leaf [16]. Thus, more characteristics could be added for the GWAS, and we will obtain more about the genetic variation in the configuration factors of sorghum yield, which is more conducive to the analysis of the genetic variation in each character and the aggregation analysis of the favorable allele variation.

#### 3.2. A More Appropriate Model or Lower Threshold for Detection

Association analysis is an important method used to analyze the genetic mechanism of complex traits [21]. Population structure analysis could reduce the impact of population classification on the division of subgroups in association analysis [38]. We found no significant loci in the number of internodes and 100-grain weight traits. The positive sites may be screened out due to the problem of model selection [17]. Additionally, the 100-grain weight trait may be a complex trait. In the further study, a higher  $p$ -value was selected for site screening. Meanwhile, 100-grain weight trait could be divided into some

other related traits, such as the grain length and width, and the significant loci may be detected. Combined with gene verification, we could also clone the genes that control the traits. Finally, some other algorithms, such as restrictive two-stage multipoint (RTM) GWASs, could be used for significant SNP identification in the 100-grain weight trait [39]. Meanwhile, more careful verification was required because of the lower threshold of RTM-GWAS. In our study, the novel QTL with higher reliability was the purpose and the RTM-GWAS could be used for further study if more QTL want to be detected.

### 3.3. *SORBI\_3008G116500* Breeding

In sorghum, researchers have identified many QTLs with multiple traits [26,28,36,40,41], and so did we. After that, the genes were cloned by GWAS, and superior alleles and markers linked to the allele were found [42]. These genes would be beneficial for molecular breeding, such as marker-assisted selection. The *AtVIP5* functioned in the heading date trait [43]. Loss of the mutants, such as the other VIP genes (for example, *VIP2/ELF7*, *VIP4*, *VIP5*, and *VIP6/ELF8*) also flower earlier because of silencing of *FLOWERING LOCUS C (FLC)* [33,44,45]. In rice, the *IPI1* gene which was annotated as *VIP2* modulates IPA1 protein levels, and the plant architecture, including plant height and branch number, were changed when overexpressing the *IPI1* [43]. In our study, we supposed that the homologue of *AtVIP5* could function not only in the heading date, but also in the spike trait. In the later stage, combined with transgenic lines or mutant lines, further validation would need to be conducted to understand the function of *SORBI\_3008G116500*, and excellent allele markers would need to be developed for molecular marker selection in sorghum breeding, especially improving the traits including heading date, plant height, spike type, and spike length. Additionally, the Chinese sorghums have formed a rich genetic base and unique characteristics have evolved as reported previously [46]. The elite haplotype of *SORBI\_3008G116500* was mostly distributed in China, validating that the sorghum germplasm in China had more potential for breeding compared with the germplasm in other countries.

## 4. Material and Method

### 4.1. Plant Materials and Environment Conditions

A total of 214 accessions comprised the varieties were used in our study. Materials No.1-No.186 were provided by Sorghum Germplasm Resource Conservation Bank of Sorghum Research Institute of Shanxi Agricultural University. Materials No.187–No.214 were collected from “The Third National Survey and Collection of Crop Germplasm Resources” Special Project for the Germplasm Resources Protection of the Ministry of Agriculture and Rural Affairs of the People’s Republic of China and were provided by the Jiangsu Academy of Agricultural Sciences. These were planted on the Liuhe base of the Jiangsu Academy of Agricultural Sciences in 2018. The test site is located at 118.63 °E and 32.48 °N. This area has a subtropical monsoon climate, with an average temperature of 15.4 °C. The average precipitation is 1073.8 mm, the average frost-free period is 225 days, and the average annual sunshine total is 1956.2 h. The standard agricultural practice was followed for plant growth and development, including irrigation, fertilizer application, and pest control. All experiments used a completely randomized block design with three replicates. The authors declare that all that permissions or licenses were obtained to collect the sorghum plant, and that all aspects of the study comply with relevant institutional, national, and international guidelines and legislation for plant ethics in the “Methods” section.

### 4.2. Genotyping

All accessions of DNA were extracted using the DNAsecure plant kit (Qiagen, Cat. No. DP320). The truseq library construction kit (FC-121-4001) was used to build the library. DNA fragments were prepared by end repair, and PloyA tail and sequencing connector were added, followed by purification and PCR amplification. The libraries were analyzed by the Bioanalyzer (Agilent, Santa Clara, CA, USA), and the PCR products were quantified



by a Qubit 3.0 fluorometer (Invitrogen, Carlsbad, CA, USA). The constructed libraries were sequenced at Illumina HiSeq Xten platform to produce 150 bp paired-end reads. The sequencing reads were extracted from the raw data; reads with an adaptor, low-quality reads, and reads with an unidentified nucleotide sequence were removed using fastp-0.20.1 (<http://opengene.org/fastp/fastp>, accessed on 3 December 2021) [47]. After that, the BWA and GATK software were used for alignment and variant calling for high-quality reads [48,49]. The reference was downloaded from NCBI website ([ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/195/GCF\\_000003195.3\\_Sorghum\\_bicolor\\_NCBIv3](ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/003/195/GCF_000003195.3_Sorghum_bicolor_NCBIv3), accessed on 3 December 2021). Moreover, we used Plink to filter the variants [50]. On the premise of minimum allele frequency (MAF)  $\geq 0.05$  and missing data  $\leq 0.2$ , 2,811,016 high-quality SNPs were obtained.

#### 4.3. Phenotypic Trait Evaluation and Data Analysis

The heading date was collected at the heading stage. At the maturity stage, the other phenotypic data of five yield-related traits were collected. For collecting plant height trait, five plants were randomly selected to be measured. Numbers of Internode were determined after stripping leaves from the stem. The data of spike type was collected according to the “Guidelines for the conduct of tests for distinctness, uniformity and stability-Sorghum (*Sorghum bicolor* (L.) Moench)” in China. However, during specific measurements, the sorghum spike phenotype has a phenotype that resides between two types. We have defined the specific type in the Supplementary File. According to the guidelines for conducting tests for distinctness, uniformity, and stability-Sorghum (*Sorghum bicolor* (L.) Moench) [GB/T 19557.15-2018], spike length was measured from the leaf scars under the spike to the top of the main spike. Lastly, mature grains of each inbred line were harvested and dried at 50 °C until the weight became constant. After drying, we collected the 100-grain weight data. The means of all phenotypic data and correlation coefficients were calculated using SAS.

#### 4.4. Population Structure, Relative Kinship, Principal Component Analysis (PCA), and Linkage Disequilibrium (LD)

Population structure (Q) was determined using Admixture software and filtered SNP markers [51]. The number of subpopulations was determined by the CV error. Relative pairwise kinship (K) was calculated using TASSEL5.0 [52]. The PCA was analyzed with SNP markers using the TASSEL 5.0, and the results were visualized using R (<https://www.r-project.org>, accessed on 25 May 2022). The LD was also calculated with Plink using the SNP data of whole genomes and each chromosome.

#### 4.5. Genome-Wide Association Study (GWAS) and QTL analysis

The GWAS for the six yield-related traits was performed for the three algorithms, including lmm, fastlmm and emmax, using GEMMA, FaST-LMM, and EMMAX programs [53–55]. The significant value was decided by 0.05/markers. The results were visualized using R package-CMplot. The LD block was defined with LDBlockShow software [56]. The gene annotation was based on Gramene (<https://www.gramene.org/>, accessed on 29 July 2022). The gene association study was conducted using TASSEL5.0 [57].

## 5. Conclusions

Overall, six traits related to yield were analyzed by three different algorithms based on natural variations in 214 landraces, and many significant loci were obtained. We also analyzed the differences of the three algorithms and the detection ability for different traits. Finally, a candidate gene was obtained that could affect multiple traits and find the elite haplotype. Our results enrich the discovery of sorghum-associated SNPs, and the discovered QTLs and gene will be helpful to sorghum breeding.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/plants12071557/s1>, Figure S1: The CV error computed by Admixture; Figure S2: The kinship of 214 landraces based on TASSEL5 analysis; Figure S3: Principal component analysis of sorghum association panel used based on the 2,811,016 SNPs. Different color of each landrace meant different subpopulation based on Admixture; Figure S4: The spike type in the study; Figure S5: The LD decay of the population (left) and each chromosome (right) based on  $R^2$ ; Figure S6: The Manhattan and Q-Q plot of six traits including heading date (HD), plant height (PL), spike type (ST), spike length (SL), number of internodes (NI), and 100-grain weight (100GW) based on the Fast-Lmm algorithm; Figure S7: The Manhattan and Q-Q plot of six traits including heading date (HD), plant height (PL), spike type (ST), spike length (SL), number of internodes (NI), and 100-grain weight (100GW) based on the Lmm algorithm; Table S1: The list of all 214 landraces; Table S2: The SNP number of chromosome.; Table S3: The result of PCA; Table S4: The all phenotype data of all 214 landraces; Table S5: The analysis of SNPs which exceeded threshold using the emmax model; Table S6: The analysis of SNPs which exceeded threshold using the fastlmm model; Table S7: The analysis of SNPs which exceeded threshold using the lmm model; Table S8: All genes in the LD block; Table S9: The all variations of three haplotypes.

**Author Contributions:** Y.Z. and Y.W. designed the experiments. X.F., X.Z. (Xiaojuan Zhang) and M.N. performed the computational analysis. Q.G., S.M., X.Z. (Xianggui Zhang), P.X., C.L., W.G. and H.W. assisted with experiments in data collection and analysis. D.L. and Q.L. participated in the design and supervised the study. Y.Z. drafted the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Special Project of Molecular Breeding Platform of Minor Crops of Shanxi Academy of Agricultural Sciences (no. YGC2019FZ5), the Open Research Fund of Shanxi Key Laboratory of Sorghum Genetic and Germplasm Innovation (No. 2018K-01), the Shanxi Key Laboratory of Minor Crops Germplasm Innovation and Molecular Breeding of Shanxi Agricultural University (No. 202105D121010-13), the Research Program Sponsored by State Key Laboratory of Integrative Sustainable Dryland Agriculture (in preparation) of Shanxi Agricultural University (202002-5), the Biological Breeding Project of Shanxi Agricultural University (YZGC062), “the Third National Survey and Collection of Crop Germplasm Resources” Special Project for Germplasm Resources Protection of the Ministry of Agriculture and Rural Affairs of the People’s Republic of China (111821301354052028), and the Shanxi Province Excellent Doctoral Work Award—Scientific Research Project (SXBKY2022071).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The resequencing data were uploaded in the NCBI database (PRJNA826146, <https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA826146>, accessed on 30 November 2022). The datasets supporting the conclusions of this article are included within the article and its additional files.

**Acknowledgments:** The authors thank the Variety Resources Research Office of the Sorghum Research Institute of Shanxi Agricultural University for providing some experimental materials.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Paterson, A.H.; Bowers, J.E.; Bruggmann, R.; Dubchak, I.; Grimwood, J.; Gundlach, H.; Haberer, G.; Hellsten, U.; Mitros, T.; Poliakov, A.; et al. The Sorghum bicolor genome and the diversification of grasses. *Nature* **2009**, *457*, 551–556. [[CrossRef](#)]
2. Ananda, G.K.; Myrans, H.; Norton, S.L.; Gleadow, R.; Furtado, A.; Henry, R.J. Wild sorghum as a promising resource for crop improvement. *Front. Plant Sci.* **2020**, *11*, 1108. [[CrossRef](#)] [[PubMed](#)]
3. Ordonio, R.; Ito, Y.; Morinaka, Y.; Sazuka, T.; Matsuoka, M. Chapter Five—Molecular Breeding of Sorghum bicolor, A Novel Energy Crop. In *International Review of Cell and Molecular Biology*; Jeon, K.W., Ed.; Academic Press: Cambridge, MA, USA, 2016; Volume 321, pp. 221–257.
4. Calviño, M.; Messing, J. Sweet sorghum as a model system for bioenergy crops. *Curr. Opin. Biotechnol.* **2012**, *23*, 323–329. [[CrossRef](#)]
5. Regassa, T.H.; Wortmann, C.S. Sweet sorghum as a bioenergy crop: Literature review. *Biomass Bioenergy* **2014**, *64*, 348–355. [[CrossRef](#)]

6. Xin, Z.; Wang, M.; Cuevas, H.E.; Chen, J.; Harrison, M.; Pugh, N.A.; Morris, G. Sorghum genetic, genomic, and breeding resources. *Planta* **2021**, *254*, 114. [[CrossRef](#)] [[PubMed](#)]
7. Jiao, X.; Lyu, Y.; Wu, X.; Li, H.; Cheng, L.; Zhang, C.; Yuan, L.; Jiang, R.; Jiang, B.; Rengel, Z.; et al. Grain production versus resource and environmental costs: Towards increasing sustainability of nutrient use in China. *J. Exp. Bot.* **2016**, *67*, 4935–4949. [[CrossRef](#)]
8. Yano, K.; Yamamoto, E.; Aya, K.; Takeuchi, H.; Lo, P.C.; Hu, L.; Yamasaki, M.; Yoshida, S.; Kitano, H.; Hirano, K.; et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* **2016**, *48*, 927. [[CrossRef](#)]
9. Hao, H.; Li, Z.; Leng, C.; Lu, C.; Luo, H.; Liu, Y.; Wu, X.; Liu, Z.; Shang, L.; Jing, H.C. Sorghum breeding in the genomic era: Opportunities and challenges. *TAG. Theor. Appl. Genetics. Theor. Und Angew. Genet.* **2021**, *134*, 1899–1924. [[CrossRef](#)]
10. Li, S.; Tian, Y.; Wu, K.; Ye, Y.; Yu, J.; Zhang, J.; Liu, Q.; Hu, M.; Li, H.; Tong, Y.; et al. Modulating plant growth-metabolism coordination for sustainable agriculture. *Nature* **2018**, *560*, 595–600. [[CrossRef](#)] [[PubMed](#)]
11. Huang, X.; Qian, Q.; Liu, Z.; Sun, H.; He, S.; Luo, D.; Xia, G.; Chu, C.; Li, J.; Fu, X. Natural variation at the DEP1 locus enhances grain yield in rice. *Nat. Genet.* **2009**, *41*, 494–497. [[CrossRef](#)]
12. Wei, X.; Xu, J.; Guo, H.; Jiang, L.; Chen, S.; Yu, C.; Zhou, Z.; Hu, P.; Zhai, H.; Wan, J. DTH8 suppresses flowering in rice, influencing plant height and yield potential simultaneously. *Plant Physiol.* **2010**, *153*, 1747–1758. [[CrossRef](#)] [[PubMed](#)]
13. Ishimaru, K.; Hirotsu, N.; Madoka, Y.; Murakami, N.; Hara, N.; Onodera, H.; Kashiwagi, T.; Ujiie, K.; Shimizu, B.-i.; Onishi, A.; et al. Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nat. Genet.* **2013**, *45*, 707–711. [[CrossRef](#)] [[PubMed](#)]
14. Chen, J.; Zhu, M.; Liu, R.; Zhang, M.; Lv, Y.; Liu, Y.; Xiao, X.; Yuan, J.; Cai, H. BIOMASS YIELD 1 regulates sorghum biomass and grain yield via the shikimate pathway. *J. Exp. Bot.* **2020**, *71*, 5506–5520. [[CrossRef](#)] [[PubMed](#)]
15. Xia, J.; Zhao, Y.; Burks, P.; Pauly, M.; Brown, P.J. A sorghum NAC gene is associated with variation in biomass properties and yield potential. *Plant Direct* **2018**, *2*, e00070. [[CrossRef](#)]
16. Zhao, J.; Mantilla Perez, M.B.; Hu, J.; Salas Fernandez, M.G. Genome-Wide Association Study for Nine Plant Architecture Traits in Sorghum. *Plant Genome* **2016**, *9*, 1–14. [[CrossRef](#)]
17. Zhu, C.; Gore, M.; Buckler, E.S.; Yu, J. Status and Prospects of Association Mapping in Plants. *Plant Genome J.* **2008**, *1*, 5. [[CrossRef](#)]
18. The Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **2007**, *447*, 661–678. [[CrossRef](#)]
19. Atwell, S.; Huang, Y.S.; Vilhjalmsen, B.J.; Willems, G.; Horton, M.; Li, Y.; Meng, D.; Platt, A.; Tarone, A.M.; Hu, T.T.; et al. Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature* **2010**, *465*, 627–631. [[CrossRef](#)]
20. Crowell, S.; Korniliev, P.; Falcao, A.; Ismail, A.; Gregorio, G.; Mezey, J.; McCouch, S. Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. *Nat. Commun.* **2016**, *7*, 10527. [[CrossRef](#)]
21. Huang, X.; Han, B. Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* **2014**, *65*, 531–551. [[CrossRef](#)]
22. Yano, K.; Morinaka, Y.; Wang, F.; Huang, P.; Takehara, S.; Hirai, T.; Ito, A.; Koketsu, E.; Kawamura, M.; Kotake, K.; et al. GWAS with principal component analysis identifies a gene comprehensively controlling rice architecture. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 21262–21267. [[CrossRef](#)]
23. Biscarini, F.; Cozzi, P.; Casella, L.; Riccardi, P.; Vattari, A.; Orasen, G.; Perrini, R.; Tacconi, G.; Tondelli, A.; Biselli, C.; et al. Genome-Wide Association Study for Traits Related to Plant and Grain Morphology, and Root Architecture in Temperate Rice Accessions. *PLoS ONE* **2016**, *11*, e0155425. [[CrossRef](#)] [[PubMed](#)]
24. Lasky, J.R.; Upadhyaya, H.D.; Ramu, P.; Deshpande, S.; Hash, C.T.; Bonnette, J.; Juenger, T.E.; Hyma, K.; Acharya, C.; Mitchell, S.E. Genome-environment associations in sorghum landraces predict adaptive traits. *Sci. Adv.* **2015**, *1*, e1400218. [[CrossRef](#)] [[PubMed](#)]
25. Chen, J.; Chopra, R.; Hayes, C.; Morris, G.; Marla, S.; Burke, J.; Xin, Z.; Burow, G. Genome-Wide Association Study of Developing Leaves' Heat Tolerance during Vegetative Growth Stages in a Sorghum Association Panel. *Plant Genome* **2017**, *10*, 1–15. [[CrossRef](#)] [[PubMed](#)]
26. Li, J.; Tang, W.; Zhang, Y.W.; Chen, K.N.; Wang, C.; Liu, Y.; Zhan, Q.; Wang, C.; Wang, S.B.; Xie, S.Q.; et al. Genome-Wide Association Studies for Five Forage Quality-Related Traits in Sorghum (*Sorghum bicolor* L.). *Front. Plant Sci.* **2018**, *9*, 1146. [[CrossRef](#)]
27. Rhodes, D.H.; Hoffmann, L., Jr.; Rooney, W.L.; Ramu, P.; Morris, G.P.; Kresovich, S. Genome-Wide Association Study of Grain Polyphenol Concentrations in Global Sorghum [*Sorghum bicolor* (L.) Moench] Germplasm. *J. Agric. Food Chem.* **2014**, *62*, 10916–10927. [[CrossRef](#)] [[PubMed](#)]
28. Kimani, W.; Zhang, L.M.; Wu, X.Y.; Hao, H.Q.; Jing, H.C. Genome-wide association study reveals that different pathways contribute to grain quality variation in sorghum (*Sorghum bicolor*). *BMC Genom.* **2020**, *21*, 112. [[CrossRef](#)]
29. Boyles, R.E.; Cooper, E.A.; Myers, M.T.; Brenton, Z.; Rauh, B.L.; Morris, G.P.; Kresovich, S. Genome-Wide Association Studies of Grain Yield Components in Diverse Sorghum Germplasm. *Plant Genome* **2016**, *9*, 1–17. [[CrossRef](#)]
30. Mace, E.S.; Singh, V.; Van Oosterom, E.J.; Hammer, G.L.; Hunt, C.H.; Jordan, D.R. QTL for nodal root angle in sorghum (*Sorghum bicolor* L. Moench) co-locate with QTL for traits associated with drought adaptation. *Theor. Appl. Genet.* **2012**, *124*, 97–109. [[CrossRef](#)]

31. Murray, S.C.; Rooney, W.L.; Hamblin, M.T.; Mitchell, S.E.; Kresovich, S. Sweet sorghum genetic diversity and association mapping for brix and height. *Plant Genome* **2009**, *2*, 48–62. [[CrossRef](#)]
32. Xue, W.; Xing, Y.; Weng, X.; Zhao, Y.; Tang, W.; Wang, L.; Zhou, H.; Yu, S.; Xu, C.; Li, X.; et al. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. *Nat. Genet.* **2008**, *40*, 761–767. [[CrossRef](#)] [[PubMed](#)]
33. Oh, S.; Zhang, H.; Ludwig, P.; van Nocker, S. A mechanism related to the yeast transcriptional regulator Paf1c is required for expression of the Arabidopsis FLC/MAF MADS box gene family. *Plant Cell* **2004**, *16*, 2940–2953. [[CrossRef](#)] [[PubMed](#)]
34. Lu, C.; Tian, Y.; Wang, S.; Su, Y.; Mao, T.; Huang, T.; Chen, Q.; Xu, Z.; Ding, Y. Phosphorylation of SPT5 by CDKD; 2 is required for VIP5 recruitment and normal flowering in Arabidopsis thaliana. *Plant Cell* **2017**, *29*, 277–291. [[CrossRef](#)]
35. Nagaraja Reddy, R.; Madhusudhana, R.; Murali Mohan, S.; Chakravarthi, D.; Mehtre, S.; Seetharama, N.; Patil, J. Mapping QTL for grain yield and other agronomic traits in post-rainy sorghum [*Sorghum bicolor* (L.) Moench]. *Theor. Appl. Genet.* **2013**, *126*, 1921–1939. [[CrossRef](#)]
36. Tao, Y.; Zhao, X.; Wang, X.; Hathorn, A.; Hunt, C.; Cruickshank, A.W.; van Oosterom, E.J.; Godwin, I.D.; Mace, E.S.; Jordan, D.R. Large-scale GWAS in sorghum reveals common genetic control of grain size among cereals. *Plant Biotechnol. J.* **2020**, *18*, 1093–1105. [[CrossRef](#)]
37. Sestili, F.; Pagliarello, R.; Zega, A.; Saletti, R.; Pucci, A.; Botticella, E.; Masci, S.; Tundo, S.; Moschetti, I.; Foti, S. Enhancing grain size in durum wheat using RNAi to knockdown GW2 genes. *Theor. Appl. Genet.* **2019**, *132*, 419–429. [[CrossRef](#)] [[PubMed](#)]
38. Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **2006**, *38*, 904–909. [[CrossRef](#)]
39. Pan, L.; He, J.; Zhao, T.; Xing, G.; Wang, Y.; Yu, D.; Chen, S.; Gai, J. Efficient QTL detection of flowering date in a soybean RIL population using the novel restricted two-stage multi-locus GWAS procedure. *TAG. Theor. Appl. Genetics. Theor. Und Angew. Genet.* **2018**, *131*, 2581–2599. [[CrossRef](#)]
40. Luo, F.; Pei, Z.; Zhao, X.; Liu, H.; Jiang, Y.; Sun, S. Genome-Wide Association Study for Plant Architecture and Bioenergy Traits in Diverse Sorghum and Sudangrass Germplasm. *Agronomy* **2020**, *10*, 1602. [[CrossRef](#)]
41. Zhang, D.; Kong, W.; Robertson, J.; Goff, V.H.; Epps, E.; Kerr, A.; Mills, G.; Cromwell, J.; Lugin, Y.; Phillips, C.; et al. Genetic analysis of inflorescence and plant height components in sorghum (Panicoidae) and comparative genetics with rice (Oryzoidae). *BMC Plant Biol.* **2015**, *15*, bbaa227. [[CrossRef](#)]
42. Springer, N.M.; Dong, H.; Zhao, H.; Xie, W.; Han, Z.; Li, G.; Yao, W.; Bai, X.; Hu, Y.; Guo, Z.; et al. A Novel Tiller Angle Gene, TAC3, together with TAC1 and D2 Largely Determine the Natural Variation of Tiller Angle in Rice Cultivars. *PLoS Genet.* **2016**, *12*, e1006412.
43. Wang, J.; Yu, H.; Xiong, G.; Lu, Z.; Jiao, Y.; Meng, X.; Liu, G.; Chen, X.; Wang, Y.; Li, J. Tissue-Specific Ubiquitination by IPA1 Interacting PROTEIN1 Modulates IPA1 Protein Levels to Regulate Plant Architecture in Rice. *Plant Cell* **2017**, *29*, 697–707. [[CrossRef](#)] [[PubMed](#)]
44. He, Y.; Doyle, M.R.; Amasino, R.M. PAF1-complex-mediated histone methylation of FLOWERING LOCUS C chromatin is required for the vernalization-responsive, winter-annual habit in Arabidopsis. *Genes Dev.* **2004**, *18*, 2774–2784. [[CrossRef](#)] [[PubMed](#)]
45. Zhang, H.; van Nocker, S. The Vernalization Independence 4 gene encodes a novel regulator of Flowering Locus C. *Plant J. Cell Mol. Biol.* **2002**, *31*, 663–673. [[CrossRef](#)]
46. Qingshan, L.; Dahlberg, J.A. Chinese sorghum genetic resources. *Econ. Bot.* **2001**, *55*, 401–425. [[CrossRef](#)]
47. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [[CrossRef](#)]
48. McKenna, A.; Hanna, M.; Banks, E.; Sivachenko, A.; Cibulskis, K.; Kernysky, A.; Garimella, K.; Altshuler, D.; Gabriel, S.; Daly, M.; et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **2010**, *20*, 1297–1303. [[CrossRef](#)]
49. Li, H.; Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **2009**, *25*, 1754–1760. [[CrossRef](#)]
50. Purcell, S.; Neale, B.; Todd-Brown, K.; Thomas, L.; Ferreira, M.A.; Bender, D.; Maller, J.; Sklar, P.; de Bakker, P.I.; Daly, M.J.; et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **2007**, *81*, 559–575. [[CrossRef](#)]
51. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [[CrossRef](#)]
52. Glaubitz, J.C.; Casstevens, T.M.; Lu, F.; Harriman, J.; Elshire, R.J.; Sun, Q.; Buckler, E.S. TASSEL-GBS: A High Capacity Genotyping by Sequencing Analysis Pipeline. *PLoS ONE* **2014**, *9*, e90346. [[CrossRef](#)] [[PubMed](#)]
53. Zhou, X.; Stephens, M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **2014**, *11*, 407–409. [[CrossRef](#)] [[PubMed](#)]
54. Lippert, C.; Listgarten, J.; Liu, Y.; Kadie, C.M.; Davidson, R.I.; Heckerman, D. FaST linear mixed models for genome-wide association studies. *Nat. Methods* **2011**, *8*, 833–835. [[CrossRef](#)] [[PubMed](#)]
55. Segura, V.; Vilhjálmsson, B.J.; Platt, A.; Korte, A.; Seren, Ü.; Long, Q.; Nordborg, M. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* **2012**, *44*, 825–830. [[CrossRef](#)]

56. Dong, S.-S.; He, W.-M.; Ji, J.-J.; Zhang, C.; Guo, Y.; Yang, T.-L. LDBlockShow: A fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief. Bioinform.* **2021**, *22*, 4. [[CrossRef](#)]
57. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Castevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.