



# Selection Signatures in Chinese Sorghum Reveals Its Unique Liquor-Making Properties

Liyi Zhang<sup>1†</sup>, Yanqing Ding<sup>1†</sup>, Jianxia Xu<sup>1</sup>, Xu Gao<sup>1</sup>, Ning Cao<sup>1</sup>, Kuiying Li<sup>2</sup>, Zhou Feng<sup>1,2</sup>, Bing Cheng<sup>1</sup>, Lengbo Zhou<sup>1</sup>, Mingjian Ren<sup>2</sup>, Xiaochun Lu<sup>3</sup>, Zhigui Bao<sup>4</sup>, Yuezhi Tao<sup>5</sup>, Zhanguo Xin<sup>6\*</sup> and Guihua Zou<sup>5\*</sup>

<sup>1</sup> Guizhou Institute of Upland Crops, Guizhou Academy of Agricultural Sciences, Guiyang, China, <sup>2</sup> College of Agriculture, Guizhou University, Guiyang, China, <sup>3</sup> Institute of Sorghum Research, Liaoning Academy of Agricultural Sciences, Shenyang, China, <sup>4</sup> Shanghai OE Biotech Co., Ltd., Shanghai, China, <sup>5</sup> Institute of Crop and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou, China, <sup>6</sup> Plant Stress and Germplasm Development Unit, Cropping Systems Research Laboratory, USDA-ARS, Lubbock, TX, United States

## OPEN ACCESS

### Edited by:

Chuanzhi Zhao,  
Shandong Academy of Agricultural  
Sciences, China

### Reviewed by:

Weihua Qiao,  
Institute of Crop Sciences (CAAS),  
China  
Feng Yu,  
Hubei University, China

### \*Correspondence:

Liyi Zhang  
lyzhang1997@hotmail.com  
Zhanguo Xin  
zhanguo.xin@usda.gov  
Guihua Zou  
zouguihuazw@163.com

† These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Plant Bioinformatics,  
a section of the journal  
Frontiers in Plant Science

Received: 19 April 2022

Accepted: 20 May 2022

Published: 09 June 2022

### Citation:

Zhang L, Ding Y, Xu J, Gao X,  
Cao N, Li K, Feng Z, Cheng B,  
Zhou L, Ren M, Lu X, Bao Z, Tao Y,  
Xin Z and Zou G (2022) Selection  
Signatures in Chinese Sorghum  
Reveals Its Unique Liquor-Making  
Properties.  
Front. Plant Sci. 13:923734.  
doi: 10.3389/fpls.2022.923734

Chinese sorghum (*S. bicolor*) has been a historically critical ingredient for brewing famous distilled liquors ever since Yuan Dynasty (749 ~ 652 years BP). Incomplete understanding of the population genetics and domestication history limits its broad applications, especially that the lack of genetics knowledge underlying liquor-brewing properties makes it difficult to establish scientific standards for sorghum breeding. To unravel the domestic history of Chinese sorghum, we re-sequenced 244 Chinese sorghum lines selected from 16 provinces. We found that Chinese sorghums formed three distinct genetic sub-structures, referred as the Northern, the Southern, and the Chishui groups, following an obviously geographic pattern. These sorghum accessions were further characterized in liquor brewing traits and identified selection footprints associated with liquor brewing efficiency. An importantly selective sweep region identified includes several homologous genes involving in grain size, pericarp thickness, and architecture of inflorescence. Our result also demonstrated that pericarp strength rather than grain size determines the ability of the grains to resist repeated cooking during brewing process. New insight into the traits beneficial to the liquor-brewing process provides both a better understanding on Chinese sorghum domestication and a guidance on breeding sorghum as a multiple use crop in China.

**Keywords:** sorghum, selective sweep, population structure, genome-wide association study (GWAS), liquor-brewing related traits

## INTRODUCTION

Sorghum [*S. bicolor* (L.) Moench,  $2n = 2x = 20$ ] is the fifth largest and wildly adaptable crop in the world. It is cultivated in more than 100 countries and regions around the world. It is mainly used for food, feed, fiber, brewing, and bioenergy. Modern sorghum, like other domesticated crops, has been improved as a consequence of long-term artificial selection (Mace et al., 2013; Smith et al., 2019). The diverse distribution and extensive morphological variation observed in multiple sorghum species imply a remarkable history of divergence and evolution. The northeast quadrant of Africa is believed to be the primary origin and diversity center of sorghum. Shortly after its early domestication around 6,000 years BP (Before present), sorghum spread to India

(Doggett, 1988; Dahlberg and Wasylikowa, 1996; Winchell et al., 2017). The Indian subcontinent is considered as the secondary center of origin of sorghum, and the earliest domesticated sorghum appeared in the late Harappan, approximately 4,000–3,700 years BP (Fuller, 2003; Boivin and Fuller, 2009). Chinese sorghum, the Kaoliangs, has a cultivation history of ~2,200 years with morphologies characteristic of bicolor type. It was inferred as the early bicolor variants introduced from northeast India (Mann et al., 1983; Doggett, 1988). It is widely believed that the domesticated sorghum was transmitted eastward from India to China either *via* land or sea routes (Fuller, 2003; Boivin and Fuller, 2009). Recently, molecular genetics studies suggested that Chinese sorghum may be of African origin (Li et al., 2010; Zhang et al., 2011). This question remains unsolved based on the current evidence from morphology, agronomy, and even genomic features.

In China, sorghum has been a key ingredient for brewing liquor for ~750 years (Shinoda, 1958; Han, 2012). Ancient Chinese, as known in the world, is the first to master the knowledge of using sorghum to brew distilled liquor (Awika and Rooney, 2004; Zhao, 2019b). Currently, sorghum is widely cultivated across agroclimatic zones from the south to the north in China, and more than 80% of sorghum grain is used to brew liquor (Chen et al., 2019). Techniques to brew distilled liquor were developed in the Tang and Song Dynasties (1,402–749 years BP), and commercialized in the Yuan Dynasty (749–652 years BP) (Joseph, 2000). Almost all Chinese famous brands of distilled liquors, such as Moutaijiu and Langjiu (Moutai-aroma liquor), Luzhoulaojiao, Wuliangye (strong-aroma liquor), and Fenjiu (light-aroma liquor), etc., are brewed with sorghum grain as a key ingredient. The first two flavored liquors are mainly from southwest region, a principal area producing liquors since ancient times. The liquor-making processes of different flavor liquors have different requirements for the physical properties of sorghum grains. The Maotai-aroma liquor produced in the Chishui River Basin in the Southwest has a complex and special brewing process, which requires nine times of steaming and boiling, eight times of fermentation, and seven times of liquor extraction, hence, the sorghum grains must be able to resist to the repeatedly steaming and stirring. The brewing process of other flavored liquors only requires one or two times of cooking to extract the liquor, which desires sorghum grains that can be cooked and broken easily (Sun, 2019). This remarkably domesticated history by human-intervention raises the question to what extent the artificial and adaptive selections for liquor-brewing shaped the local Chinese sorghum population.

High-throughput resequencing technology, combined with comprehensive statistical analysis methods, provides a powerful tool to uncover the mysteries of plant domestication. Based on deep-coverage of whole-genome sequence of accessions from various regions, the origin and domestication history of several crops have recently been uncovered, including ancient sorghum (*S. bicolor ssp. bicolor* (L.) Moench.) (Smith et al., 2019), Tibetan barley (*Hordeum vulgare* L., qingke) (Zeng et al., 2018), castor (*Ricinus communis*) (Fan et al., 2019; Xu et al., 2021), pear (*Pyrus*) (Wu et al., 2018), chickpea (*Cicer arietinum* L.) (Varshney et al., 2019), etc. Therefore, we investigated the data from

whole genome resequencing of 333 sorghum accessions (244 sequenced in this study plus 89 genomes available publicly) to further explore the domestication history and genome-wide selective footprints on Chinese sorghum population. In addition, we described a genome wide map of SNP variation and traced patterns of sorghum diffusion to diverse agroclimatic regions. Our study demonstrated that Chinese sorghum was first introduced to south China from Indian and gradually spread to the north China. Furthermore, we identified genes underlying natural variation in traits related to baijiu-making process.

## MATERIALS AND METHODS

### Plant Material and Phenotyping

A worldwide sorghum collection of 333 accessions was used in this study, of which a panel of 244 was mainly composed of Chinese sorghum from the Center for Crop Germplasm Resources, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences (Beijing, China), and Institute of Upland Crops, Guizhou Academy of Agricultural Science (Guiyang, China). Another panel of 89 accessions included 38 wild sorghum and 51 landraces/cultivars. Wild sorghums were mostly from Africa, while cultivated sorghums were from Africa and Asia, maintained in the Institute of Botany, Chinese Academy of Sciences (Beijing, China).

From 2018 to 2020, the panel of 244 accessions were grown in Guiyang, Guizhou Province; Lingshui and Ledong, Hainan Province; and Hangzhou, Zhejiang Province in China. Five traits related with brewing properties were thousand grain weight (TGW), grain length (GL), grain width (GW), pericarp thickness (PT) and testa thickness (TT). TGW, GL, and GW were measured by a Crop Grain Appearance Quality Scanning Machine (SC-E, Wanshen Technology Company, Hangzhou, China). PT and TT were measured by a Digital Microscope (VHX-7000, KEYENCE Technology Company, Shanghai, China).

### DNA Extraction and Sequencing

Genomic DNAs of 244 accessions were extracted following the cetyltrimethylammonium bromide (CTAB) method and quantified by Qubit® 2.0 fluorescent meter (Invitrogen, Carlsbad, United States). The quality of the DNA was determined by electrophoresis in 0.8% agarose gel running at 100 V for 40 min. The DNA fragments around 350 bp were randomly generated by Covaris ultrasonic crushing apparatus. The constructed library was used for paired-end (PE150) sequencing on Illumina HiSeq 4000 sequencing platform by Beijing Novogene technology co., Ltd. (Beijing, China). The 89 public sequencing data for African and Asian sorghums were obtained from SorGSD (Zhang et al., 2018)<sup>1</sup>.

### Reads Mapping and Variants Calling

The raw paired-end reads were trimmed and filtered with a sliding window of size 4, with average Phred score scale of 20 within the window using fastp1 (version: 0.20.0)

<sup>1</sup><http://bigd.big.ac.cn/sorgsd>

(Chen et al., 2018). The clean reads of 333 accessions were mapped to the *Sorghum bicolor* genome (Phytozome v12<sup>2</sup>) (Goodstein et al., 2012) using bwa-mem (version: 0.7-17) (Li and Durbin, 2009) with default parameters. After alignment, Picard tools (version: 2.18.17<sup>3</sup>) were used to remove PCR duplicates according to the mapping coordinates.

The variation detection followed the best practice workflow recommended by Genome Analysis Toolkit (GATK4, version 3.8.1) (McKenna et al., 2010). In brief, the variants were called for each accession by the GATK HaplotypeCaller. A joint genotyping step for comprehensive variations union was performed on the gVCF files. Two subsets of the sorghum SNPs were defined using the following filtering criteria: (1) the basic set of 3,298,433 SNPs were created with exclusion of non bi-allelic, > 20% missing calls and MAF < 5%; (2) a core SNP set of 526,946 SNPs derived from the basic SNP set using a two-step linkage disequilibrium pruning procedure with PLINK (version: 1.9) (Purcell et al., 2007) with parameters “-indep-pairwise 50 5 0.5.”

SNPs and Indels were annotated according to the BTx623 genome using the SNPeff6 (version: 4.3T) (Cingolani et al., 2012). SNP and InDel densities across each chromosome were counted with 500 kb sliding window using VCFtools (version: 0.1.17) software (Danecek et al., 2011).

## Phylogenetic and Population Analyses

A phylogenetic tree was constructed from the core SNP set by using the neighbor-joining method in the program PHYLIP (version: 3.697<sup>4</sup>), and IBS distance matrices were calculated using PLINK. The resulting phylogenetic tree was visualized using the online tool iTOL (version: 4.0) (Letunic and Bork, 2019).

Principal Components Analysis (PCA) was performed with the smartPCA program from EIGENSOFT package (version: 6.1.4) (Price et al., 2006) with the core SNP set, and the first three eigenvectors were plotted by scatterplot3d R package. Population structure was inferred using the fastSTRUCTURE (version: 1.0) program (Raj et al., 2014). To explore the convergence of individuals, we predefined the number of genetic clusters *K* from 2 to 12 and ran the cross-validation error (CV) procedure. fastSTRUCTURE was then run again on the whole core SNP set 10 times with varying random seeds; the Q-matrices were aligned using pong13 software and clustered based on similarity. Then, the matrices belonging to the largest cluster were averaged to produce the final matrix of admixture proportions.

Global and chromosomal *r*<sup>2</sup> for each sorghum group were evaluated using PopLDdecay (version: v3.41<sup>5</sup>) (Zhang et al., 2019) with default setting.

## Genome Scanning for Selective Sweep Signals

We performed a genetic differentiation (*Fst*), nucleotide diversity ( $\pi$ ) and Tajima'D based cross approach to investigate the

selection signals across the whole genome. A 10 kb sliding window with 10 kb step approach was applied to quantify *Fst*,  $\pi$  and Tajima'D by using VCFtools software (Danecek et al., 2011). The candidates that meet both top 5% of the three tests were selected as selective signals.

## Genome-Wide Association Study

The association analysis was performed with the GLM and Blink statistical methods implemented in Genome Association and Prediction Integrated Tool package (GAPIT version: 3.0)<sup>6</sup>. The first three PCs derived from whole-genome SNPs were used as fixed effects in the mixed model to correct for stratification. The random effect was estimated from the groups clustered based on the kinship among all accessions. We defined the whole-genome significant cutoff with the adjusted Bonferroni test threshold, which was set as  $P = -\log_{10}(0.05/2,015,850) = 7.6$ . LD analysis and inference of the haplo-type blocks containing peak SNPs to document the potential candidate genes responsible for traits were performed in LDBlockShow (version: 1.39) (Dong et al., 2021).

## Sequence Analysis of the Candidate Gene

The candidate gene (Sobic.002G228600) from 20 accessions with extreme phenotypes were amplified and sequenced by Sanger sequencing technology in Platinum Biotechnology Co., Ltd. (Shanghai, China). The raw sequences were assembled using DNASTAR Lasergene (version: 7.10) and alignments for the sequence dataset were performed by MAFFT (version: 7.48) to identify the variations. The ExpASY translate tool<sup>7</sup> was used to convert nucleotide sequence into amino acid sequence, the sequencing results are shown in **Supplementary Figure 4**. The primers1 information: 5'-TTGGACAGTGTGGACTCTC-3' and 5'-CAGGAAGTAGGGATGGATCA-3', the primers2 information: 5'-TCGTCGTGTCGGTGAATAC-3' and 5'-TGTAACCTCAGCACGCAACAA-3'.

## RNA-Seq and Data Analysis

Two sorghum cultivars, “654” and “LTR108” were chosen for RNA sequencing (RNA-seq). Young panicles samples were collected at the S1, S2, and S3 stages, which corresponded to less-than 5 cm in length, 10–15 cm in length, and 5 days after heading, respectively. Total RNA was isolated using a protocol reported previously (Zou et al., 2020). A total of 1  $\mu$ g RNA was used for library construction with an NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, United States) according to the manufacturer's instructions. RNA sequencing was performed on NovaSeq 6000 platform. Clean reads were mapped to the sorghum reference genomes V3.1.1 on Phytozome database<sup>8</sup> using Hisat2 (version: 2.2.1). The number of reads mapped to each gene was counted using HTSeq software (version: 0.9.1). Gene expression based on fragments per kilo-base of exon per million fragments mapped (FPKM) was next obtained

<sup>2</sup><https://genome.jgi.doe.gov/portal/pages/dynamicOrganismDownload.jsf?organism=Phytozome>

<sup>3</sup><http://broadinstitute.github.io/picard/>

<sup>4</sup><http://evolution.genetics.washington.edu/phylip.html>

<sup>5</sup><https://github.com/BGI-shenzhen/PopLDdecay>

<sup>6</sup><https://www.maizegenetics.net/gapit>

<sup>7</sup><http://web.expasy.org/translate/>

<sup>8</sup><https://phytozome.jgi.doe.gov>

through Cufflinks (version: 2.2.1). Each sample consisted of three biological replicates.

## Experiment of Resistance to Cooking

We performed the cooking experiment on 16 glutinous sorghums whose grain size and shape, thickness of pericarp and testa were distinctly dissimilar. A volume of the grains ( $V_0$ ) with weight of 10 g of sorghum grains were measured by the drainage method in a 150 mL beaker. Beakers of glutinous and non-glutinous sorghum varieties were filled with 85°C water on standing in a thermostat for 8 and 10 h, respectively. Subsequently, the grain volume ( $V_t$ ) was measured once again without the supernatant, and the expansion rate was thus obtained [Expansion rate =  $(V_t - V_0)/V_0 \times 100\%$ ]. The crack rate was observed and counted from 100 random grains (Liu et al., 2012).

## RESULTS

### Whole Genome Resequencing of Chinese Sorghum

To investigate the population structure of Chinese sorghum accessions, we sequenced 244 sorghum landraces/cultivars, which were collected extensively from 16 sorghum planting provinces across China (151 from North China, 91 from South China, and 2 from abroad). To better address the domestication history of sorghum in China and identify the different genomic regions under domestication, we also supplemented our sequencing data with 89 publicly available sorghum genome sequences, including one accession of *Sorghum propinquum*, a wild relative of sorghum (the same subgenera with *S. bicolor*), 37 wild sorghum mainly from Africa, and 51 cultivated sorghum from Africa and Asia (Zhang et al., 2018). In total, a collection of 333 sorghum accessions was analyzed in the present study (Figure 1A and Supplementary Table 1).

The whole-genome resequencing generated a total of  $\sim 12.04$  G paired-end reads 150 bp in length, with an average sequencing depth of  $\sim 7.04 \times$  and an average genome coverage of  $\sim 86.8\%$  (Supplementary Table 2). After mapping to the sorghum reference genome BTx623 (Paterson et al., 2009; McCormick et al., 2018) and single nucleotide polymorphism (SNP) calling, we obtained 3,298,433 high-quality SNPs and 429,794 InDels (insertions-deletions) from the 333 sorghum accessions. Among these variations, 203,278 (4.2%) SNPs were in coding regions, with a non-synonymous-to-synonymous substitution ratio of 1 (86,150 and 87,333 for non-synonymous and synonymous, respectively).

### Genetic Structuration, Phylogenetic and Principal Component Analysis

The population structure of the whole data set ( $N = 333$ ) was analyzed by *fastSTRUCTURE* with  $K$  ranging from 2 to 12 (Figure 1B). The barplots were inspected visually to identify the well delimited clusters that could be biologically relevant. Although the optimal number of clusters is inferred by the  $K$  selector as 5, the finest biologically relevant population structure

was identified at  $K = 10$ . The number of main clusters remained as six, including the wild relatives, cultivated sorghum in Africa, Southern Asia, Southern China, Northern China, and Chishui in Southwest China, which were apparently consistent with the evolutionary history of sorghum and corresponded to the species and/or geographical regions (Fuller, 2003; Boivin and Fuller, 2009). We therefore considered these six clusters as the most relevant genetic structure and estimated the sorghum individual membership at  $K = 10$ .

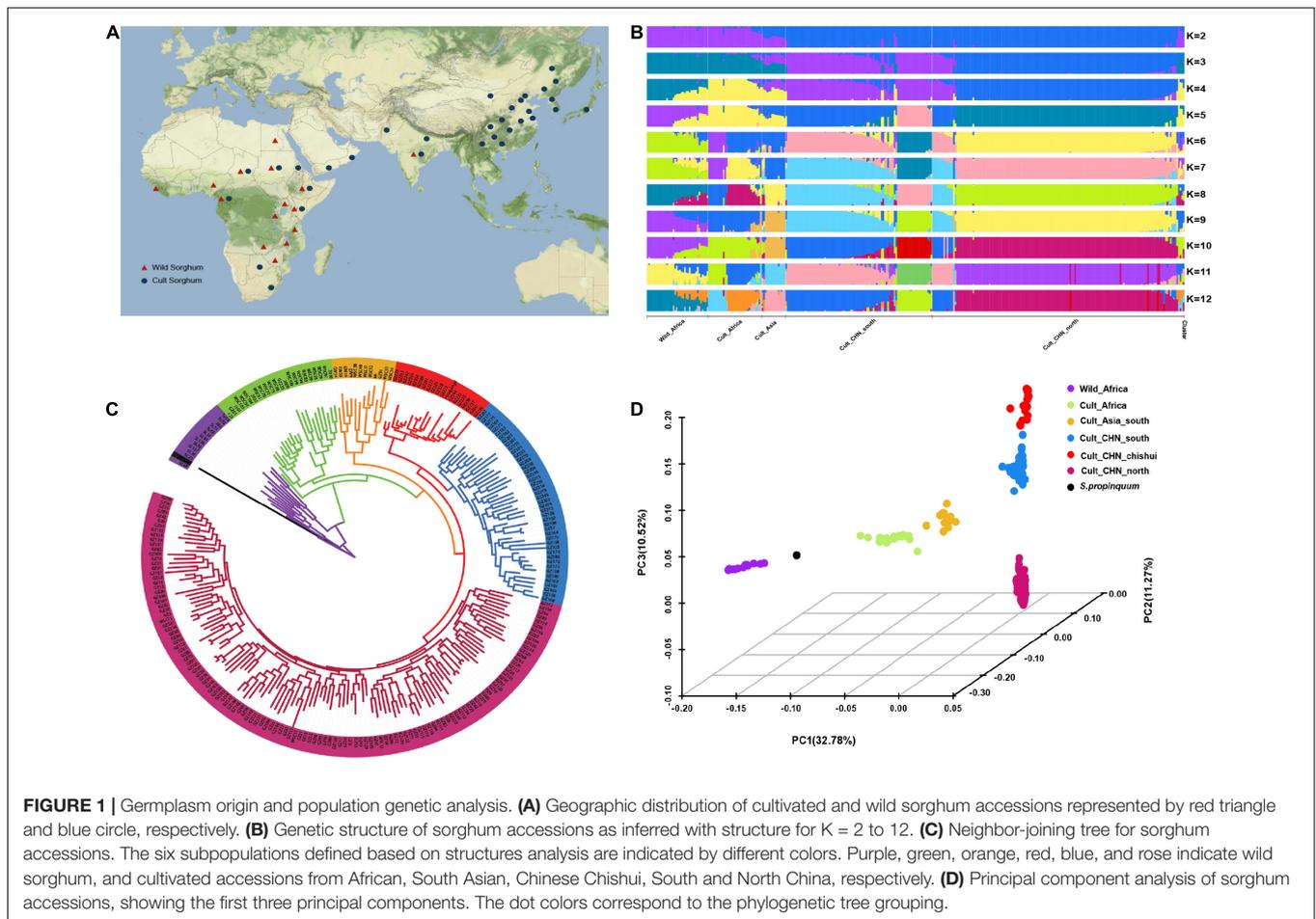
We considered hereafter a genotype to be unequivocally assigned to a population when its assignment probability was  $\geq 95\%$  to one of the six genetic clusters at  $K = 10$ , which was the case for 71.5% of individuals ( $N = 238$ ). It is important to set assignment thresholds in order to distinguish recent hybridization (or recent gene flow) from ancient gene flow, and the threshold we set appeared to be optimal (Groppi et al., 2021). Our maximum likelihood based phylogenetic dendrogram clearly distinguished six clusters (Figure 1C). The wild sorghum (purple) and African cultivated lines (green) split into two clades. The latter includes two forms of sorghums in north and south Africa, which was closely followed by South Asian sorghum (orange) and Chinese sorghum. Among Chinese accessions, sorghum from the Chishui (red) had a closer relationship with South Asian sorghum, while Northern sorghum (rose) was clearly separated from South Asian sorghum.

The above classification based on *fastSTRUCTURE* was also supported by principal component analysis. A total of 53.57% of the genetic variation was explained by the first principal component analysis (Figure 1D). The wild sorghum (purple), African cultivated sorghum (green), South Asian cultivated sorghum (orange) and Chinese accessions were well distinguished along the  $PC_1$  axis, while the sorghum of Northern China (rose), Southern (blue) and Chishui (red) were obviously separated by  $PC_3$  axis. South China sorghum were observed closer to South Asian sorghum than to the North China sorghum, which was supported by the phylogenetic relationship.

### Genetic Diversity, Linkage Disequilibrium and Selective Sweep on Sorghum Collections

Genetic divergence among the inferred six genetic groups of sorghum was firstly estimated by the genetic distance (Fixation index values,  $F_{st}$ ) across the genome (Figure 2A). The wild African sorghum had a lowest  $F_{st}$  (0.43) compared with the cultivated African sorghum, but the highest  $F_{st}$  (0.75) against the Chinese Chishui sorghum. However, the cultivated African sorghum showed lower  $F_{st}$  (0.48) with the Chinese sorghum in the south than that (0.53) of the north. These results were consistent with the phylogenetic and structure analyses.

We further implemented the mean nucleotide diversity ( $\pi$ ) and Tajima's  $D$  to address the genetic diversity of the collections (Figures 2A,B). The  $\pi$  in African and Asian cultivated sorghum genomes were  $1.14 \times 10^{-3}$  and  $1.03 \times 10^{-3}$ , respectively. Low levels of  $\pi$  diversity were found in Chinese sorghum from the Northern, the Southern, and the Chishui sub-clusters ( $7.21 \times 10^{-4}$ ,  $7.18 \times 10^{-4}$ , and  $5.52 \times 10^{-4}$ , respectively). More



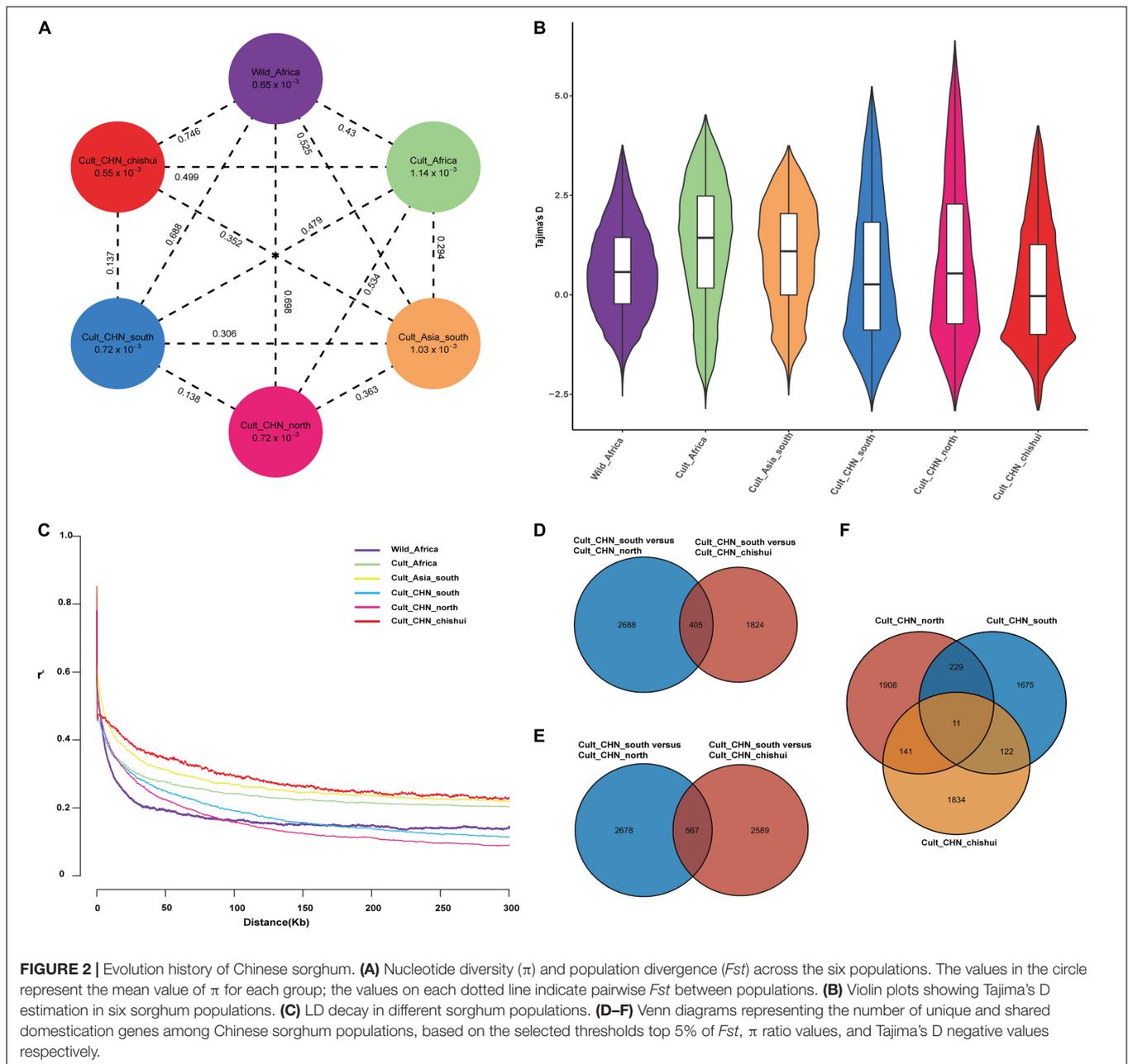
than half of Tajima's  $D$  values observed from Chishui sorghum genome were negative.

In our collections, the LD extended to background level at  $\sim 50$  kb in wild African sorghum, and at  $\sim 100$ ,  $\sim 110$ , and  $\sim 200$  kb for African, South-Asian and Chinese cultivated sorghum, respectively (Figure 2C). The LD varied on different chromosomes and in different sorghum populations (Supplementary Figure 1). The wild African sorghum possessed rapid LD declines on both chromosome 6 and 10. The Southern sorghum appeared to have higher LD level than the Northern sorghum, and the Chishui sorghum showed the lowest LD decay on chromosome 5.

To identify the potential domestication signatures in Chinese sorghum, we summarized the selective sweeps within 10-kb genomic intervals and scored significant selection by top 5% outliers in  $F_{st}$ ,  $\pi$  ratio and Tajima's  $D$  tests. Pairwise  $F_{st}$  statistics, respectively, revealed 3,093 and 2,229 candidate genes from the sweeps between Southern and Northern sorghum, and between Southern and Chishui sorghum, of which 405 candidate genes were shared (Figure 2D and Supplementary Figure 2). Meanwhile, candidate genes behind the selection signatures were identified as 3,245 by  $\pi_{Southern}/\pi_{Northern}$  ratio and as 3,156 by  $\pi_{Southern}/\pi_{Chishui}$  ratio, of which 567 candidate genes overlapped in the genomic regions (Figure 2E and

Supplementary Figure 2). Comparing the overlapping candidate genes between  $F_{st}$  and  $\pi$  ratio methods, 1,050 and 285 identical genes were found between Northern and Southern sorghums and Southern and Chishui sorghums, respectively. Additionally, the negative Tajima's  $D$  indicator was used to investigate the signals of the positive selections. We identified 2,289, 2,036, and 2,108 candidate genes for Southern, Northern and Chishui sorghum, respectively. Sorghum accessions from the Southern China shared 229 selective candidate genes against those from the Northern China and shared 122 candidate genes with Chishui sorghum. Totally, eleven candidate genes behind positive selection were in common among the three Chinese sorghum groups (Figure 2F and Supplementary Figure 3).

To understand putative functions of the strong selective sweep signals among Chinese populations, we investigated the annotations for the top 1% of the domestication genes identified by three methods. Only about 20% of these domesticated genes were matched to putative functions due to the lack of annotation. Overall, around half of the important genes are related to biotic and abiotic stress resistance, such as salt tolerance, cadmium tolerance, disease resistance, etc., while approximately 30% orthologous genes are associated with yield, quality, and plant type. The functions of important domestication genes did not differ significantly between populations based on  $F_{st}$



and  $\pi$  ratio, whereas the functions of important domesticated genes were quite distinct in different groups based on Tajima's D metric. There were more key domestication genes related to panicle structure, tillering, height, and yield in southern and Chishui populations compared with those in northern population (Supplementary Table 3).

## Impact of Selection for Traits Associated With Liquor-Brewing on Sorghum Genome

Since more than 30 processes are required to brew Maotai-flavor liquor in the Chishui River Basin in Southwest China,

the sorghum grains are required to have good cooking resistance. GWAS was thence performed to determine signals for the domestication traits-related brewing properties, based on surveying grain physical properties TGW, GL, GW, PT, and TT of the 244 Chinese landraces/cultivars in six environments from 2018 to 2020 (Supplementary Table 4). Totally, 66 significant SNP loci were identified on the 15 genome regions (Supplementary Table 5).

Two significant GWAS signals affecting both TGW and GW between Southern and Northern population during domestication were mapped in the selection sweeps associated with traits for liquor brewing were identified in this study (Supplementary Tables 3, 5). One important locus at the

position 47,459,472 bp on chromosome 9 fell into a LD block including several genes, of which the gene *Sobic.009G124200* (47,743,018 bp) is one of the strongest domestication genes based on *Fst* metric. This gene is an ortholog of the rice *SMALL ORGAN SIZE1* (*SMOS1*), which encodes an AP2-Type transcription factor acting as an auxin-dependent regulator for cell expansion during grain development (Aya et al., 2014). Another significant signal for TGW and GW was located on chromosome 10 at 11,199,795 bp position and underwent selection in both Southern and Chishui population during domestication based on Tajima's *D* metric (Figures 3A,B). One gene *Sobic.010G111200* (11,197,868 bp) fell within LD block of significant loci (Figure 3C) and is orthologous to rice *OsGASR7/GW6* that encodes a GA-regulated GAST family protein and positively regulates grain width and weight (Shi et al., 2020). As expected, Chishui and Southern sorghums have smaller grains than Northern sorghum (Figure 3D). The mRNA expression of *Sobic.010G111200* in developing panicles of less than 5-cm-long was significantly higher in cultivar LTR108 (larger grain) than in cultivar 654 (smaller grain), while mRNA levels in LTR108 dropped dramatically with panicle development and was even lower than the level of 654 after heading (Figure 3E).

We also identified two GWAS signals for TT and PT (Supplementary Table 5). A GWAS signal associated with TT was detected on chromosome 2 at 8,300,524 bp position, which is clearly distinct from the previously reported *B2* locus (60.54–60.59 Mb) that control the presence of testa (Dufour et al., 1997; Rami et al., 1998; Mace et al., 2019). In addition, a significant signal was identified for pericarp thickness on chromosome 2 (61.99 Mb) that is close to the previously reported *Z* gene locus (57.03–59.10 Mb), which is related to the thickness of pericarp and pearly pericarp (Tao et al., 1998; Boivin et al., 1999; Mace et al., 2019). The signal was located in a large genomic region (140 kb, 61.96–62.10 Mb) of the strongly selective sweep during domestication in Chishui population based on Tajima's *D* metric (Figures 4A,B). The phenotypic distribution indicated that the pericarp of Chishui and southern accession was thinner than that of northern accession (Figures 4D,E). The significant SNP loci fell within a high LD block including two genes (*Sobic.002G228600* and *Sobic.002G228700*) (Figure 4C). By comparing the sequence mutations of *Sobic.002G228600* gene between extreme phenotypic varieties, we detected an insertion and deletion of "CATATTACAATCC" in the 5'-UTR in accessions with thin and thick pericarp, respectively (Figure 4F and Supplementary Figure 4).

## Relationship Between Baijiu-Brewing Process and the Physical Properties of Grains

Chishui sorghum has consistently thinner pericarp (exocarp and pericarp) than the other two Chinese populations (Figure 3E). Correlation analysis revealed that sorghum pericarp thickness showed significant positive association with TGW, GL and GW in six environments, with the average Pearson coefficient (*R* value) of 0.32, 0.24 and 0.26 ( $P < 0.01$ ), respectively

(Supplementary Table 6). The cooking experiment was performed to investigate the relationship between resistance to cooking with grain size and thickness of pericarp and testa. For glutinous sorghum, grain with thinner pericarp and testa generally had lower the rate of expansion and crack and presented better resistance to cooking (Table 1).

## DISCUSSION

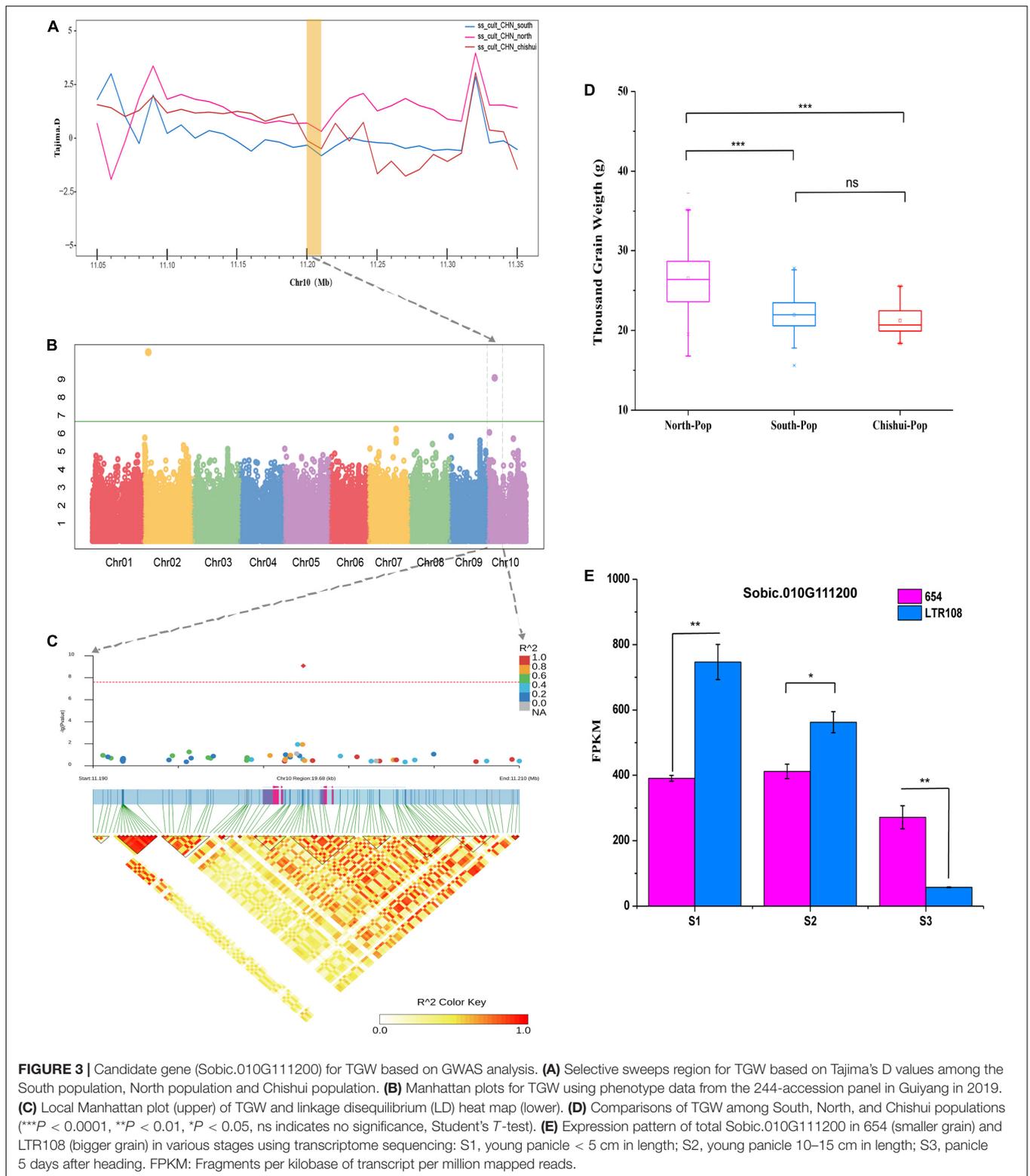
### The Origin of Chinese Sorghum

A comprehensive understanding of the origin and domestication history of crops will help us better manage and use it. In this study, we conducted substantial investigations on population genomics and evolutionary history of Chinese sorghum based on resequencing of 333 diverse accessions. Our population genetics analyses revealed that sorghum accessions from South China were more closely related to wild relatives and landraces/varieties from African and South Asian. Similar patterns were observed in Chinese castor originated in Africa (Fan et al., 2019) and Chinese cotton originated in India (Du et al., 2018), Southern varieties are more closely related to wild relatives than Northern varieties. As the local residents in the southwestern border of China have historically maintained close contact with the people of the Indochina Peninsula and the Indian subcontinent (You, 1989). This is also the origin of the ancient Chinese name of sorghum, "ShuShu," which means millet-like crop grown in the Kingdom of Shu, located in the southwestern region of China (Zhao, 2019a). We hence believe that sorghum may have been introduced from India to the southwestern region of China and subsequently spread into middle and northern China, supporting the hypothesis that sorghum entered China through the Southwest Silk Road (Zhao et al., 2019). Due to the limited sampling size (38 African and 21 Asian cultivated accessions) used in this study, this conclusion is considered tentative and needs further validation with larger sample sizes.

### Linkage Disequilibrium and Selective Sweep on Sorghum Collections

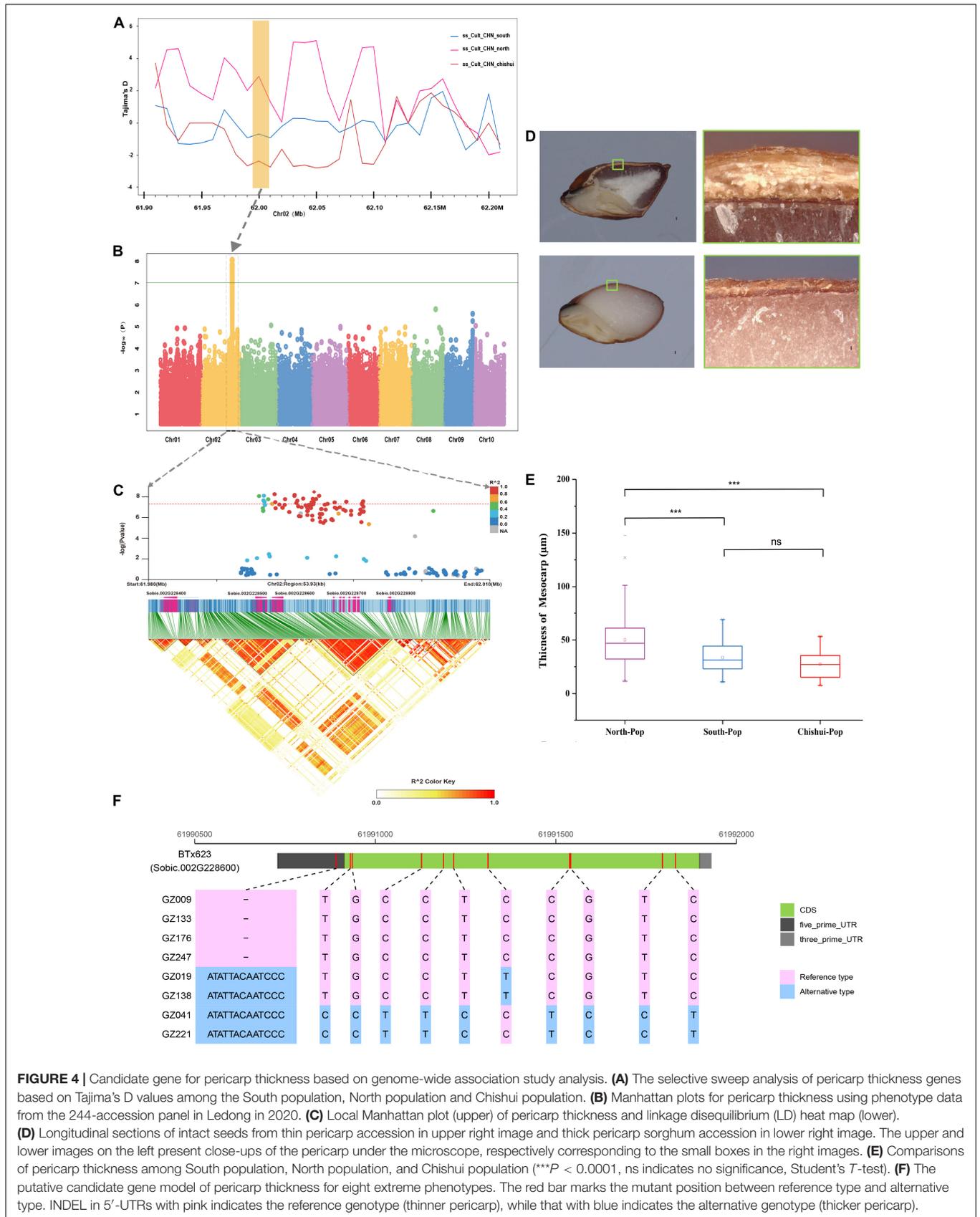
Sorghum, as an often cross-pollinated crop, has a lower linkage disequilibrium (LD) decay rate than self-pollinated rice (120 kb and 160 kb for indica and japonica, respectively) (Xu et al., 2011), and much higher than open-pollinated maize (~30 kb) (Hufford et al., 2012). The LD from wild sorghum to Chinese cultivated accession extended to background level from ~50 to ~200 kb in our collection. This is consistent with LD estimates in sorghum by using GBS and WGS techniques (Mace et al., 2013; Morris et al., 2013). In addition, chromosomes with the slowest decay rate were different in different Chinese sorghum populations, and similar phenomenon was also observed in bread wheat (Hao et al., 2020). The possible reasons are that local agroecological environments and human preference, or some genes for important agronomic traits clustering in the same segment, which causes each chromosome to experience different selection pressures during domestication in the different populations.

Our research also revealed that more genes have been subject to domestication between southern and northern populations



than southern and Chishui populations, suggesting that the number of selective sweep signals varies between different populations. Likewise, around half of the important genes is related to biotic and abiotic stress resistance, while approximately

30% orthologous genes are associated with yield, quality, and plant height. This confirms that crop domestication is the result of a synergistic effect of natural and artificial selection (Smith et al., 2019).



**FIGURE 4 |** Candidate gene for pericarp thickness based on genome-wide association study analysis. **(A)** The selective sweep analysis of pericarp thickness genes based on Tajima's D values among the South population, North population and Chishui population. **(B)** Manhattan plots for pericarp thickness using phenotype data from the 244-accession panel in Ledong in 2020. **(C)** Local Manhattan plot (upper) of pericarp thickness and linkage disequilibrium (LD) heat map (lower). **(D)** Longitudinal sections of intact seeds from thin pericarp accession in upper right image and thick pericarp sorghum accession in lower right image. The upper and lower images on the left present close-ups of the pericarp under the microscope, respectively corresponding to the small boxes in the right images. **(E)** Comparisons of pericarp thickness among South population, North population, and Chishui population ( $***P < 0.0001$ , ns indicates no significance, Student's *T*-test). **(F)** The putative candidate gene model of pericarp thickness for eight extreme phenotypes. The red bar marks the mutant position between reference type and alternative type. INDEL in 5'-UTRs with pink indicates the reference genotype (thinner pericarp), while that with blue indicates the alternative genotype (thicker pericarp).

**TABLE 1** | Results of cooking experiment for glutinous sorghum.

Name	Pericarp thickness	Testa thickness	Total thickness	Thousand grain weight	Expansion rate $\pm$ SD	Crack rate $\pm$ SD
GZ243	27.82	13.03	40.85	16.57	115.18% $\pm$ 9.88%	83.00% $\pm$ 6.67%
GZ060	36.49	14.25	50.74	20.21	114.78% $\pm$ 4.94%	81.00% $\pm$ 3.33%
GZ241	45.89	12.64	58.52	22.90	111.43% $\pm$ 5.37%	90.00% $\pm$ 0%
GZ136	48.20	17.37	65.57	23.81	91.67% $\pm$ 10.02%	77.50% $\pm$ 2.33%
GZ162	50.05	18.87	68.91	25.77	138.89% $\pm$ 7.84%	96.50% $\pm$ 0%
GZ093	50.20	4.45	54.64	29.13	117.1% $\pm$ 7.26%	73.50% $\pm$ 4.43%
GZ150	56.84	29.28	86.12	30.62	92.11% $\pm$ 2.34%	9.00% $\pm$ 0.67%
GZ210	72.35	7.75	80.10	19.18	111.43% $\pm$ 2.61%	81.50% $\pm$ 1.67%
GZ234	76.05	16.47	92.52	21.86	102% $\pm$ 2.61%	89.50% $\pm$ 2.33%
GZ202	84.60	23.13	107.73	16.23	126.87% $\pm$ 10.75%	96.50% $\pm$ 2.33%
GZ187	88.81	17.53	106.34	22.29	169.44% $\pm$ 5.74%	99.00% $\pm$ 0.67%
GZ192	88.90	14.59	103.49	17.46	169.09% $\pm$ 9.30%	86.50% $\pm$ 4.33%
GZ104	97.01	16.44	113.45	28.63	146.02% $\pm$ 1.16%	93.00% $\pm$ 0%
GZ220	99.47	16.32	115.78	26.26	133.33% $\pm$ 0%	98.00% $\pm$ 3.33%
GZ061	135.85	34.45	170.30	21.81	190.71% $\pm$ 10.02%	100.00% $\pm$ 0%
GZ123	157.34	0.00	157.34	26.56	151.08% $\pm$ 5.16%	94.50% $\pm$ 1.00%

## Domestication of Sorghum Grain Physical Properties Related to Liquor Brewing Process

The Southwest region in China, especially in the Chishui River Basin between Guizhou and Sichuan province, is the birthplace and main producing area of Maotai-aroma liquor. It is generally believed that the glutinous sorghum with features of small grain and thick hull can withstand 30 processes and several rounds of steaming and boiling in brewing Maoutai's. Phenotype data showed Chishui and Southern sorghums have smaller grains than Northern sorghum, and the GWAS identified two important domestication genes *Sobic.009G124200* and *Sobic.010G111200* for TGW and GW. The two genes are an ortholog of the *SMALL ORGAN SIZE1 (SMOS1)* and *OsGASR7/GW6* controlling grain size in rice (Aya et al., 2014; Shi et al., 2020). The varieties possessing longer and looser panicles with pendulous branches are more suitable for the hot and humid climate in southern China in summer due to their resistance to mold and insects at the stage of maturity, whereas sorghum panicle length is negatively correlated with grain size (Tao et al., 2021), indicating that natural selection for panicle length genes in southern and Chishui populations is likely to cause selective sweep of genes controlling grain size.

Grain pericarp thickness failed to catch research's attention, probably because they have no obvious relationship with traits such as yield, quality, and resistance to biotic and abiotic stresses. The *Z* gene has not been extensively studied since it was firstly reported in 1930s (Ayyangar and Ayyar, 1938), which is often used as a morphological marker to construct genetic maps in sorghum (Tao et al., 1998; Boivin et al., 1999; Mace et al., 2019). Varieties with light-colored grains are easy to determine whether pericarp is pearly, but those with dark-colored grains are difficult to determine whether pericarp is pearly, which might lead to an inaccurate genetic mapping of *Z* gene on chromosome. Based on the observation under a digital microscope, our GWAS analysis identified an important loci of pericarp thickness that is  $\sim$ 2 Mb away from the previously reported *Z* gene locus (Mace et al., 2019). The possible candidate

gene *Sobic.002G228600* is an ortholog of *GRMZM5G898880* involving in striga resistance in maize (Adewale et al., 2020), and a QTL for striga resistance was reported in nearby areas (60.58–60.80 Mb) in a previous sorghum study (Haussmann et al., 2004). *Sobic.002G228600* gene showed an insertion and deletion of "CATATTACAATCC" in the 5'-UTR in accessions with thin and thick pericarp. 5'-UTR contains key elements of translational regulation, such as structural motifs and upstream open reading frames (uORFs). By controlling the selection of translation initiation sites (TISs), many sequence elements in 5'-UTR contribute to mRNA translatability (Hinnebusch et al., 2016). For example, the significant InDels in the 5'-UTR alter the expression of *BRB* (Big Root Biomass) in sesame (Dossa et al., 2021), and an 8-bp InDel in the 5'-UTR of *SibHLH59* regulates ascorbate biosynthesis in tomato (Ye et al., 2019).

We found that *Z* gene is located in a  $\sim$ 140 kb (61.96–62.10 Mb) regions on chromosome 2 that underwent the strong artificial selection in Chishui population. In addition to grain pericarp thickness, this significant region is likely to be related to grain size, panicle shape, stem morphology and disease resistance. Several genes in the vicinity of *Z* gene locus, such as *Sobic.002G228800*, *Sobic.002G228900*, *Sobic.002G229000*, and *Sobic.002G229100*, are homologous to the genes controlling panicle architecture and seed size in Arabidopsis, rice, and maize. The gene *Sobic.002G228700* encodes an ortholog of *PATATIN-RELATED PHOSPHOLIPASE A (DEP3)*, which affects the architecture of inflorescence in rice (Qiao et al., 2011), as well as the cellulose content and cell elongation in Arabidopsis (Li M. et al., 2011). The gene *Sobic.002G228800* encodes an ortholog of *C-TERMINALLY ENCODED PEPTIDE (OsCEP6.1)* in rice, which regulates the development of panicle type and grain size by changing cell size without altering cell number (Sui et al., 2016). Two genes (*Sobic.002G229000* and *Sobic.002G229100*) encode not only an ortholog of a *PUTATIVE SERINE CARBOXYPEPTIDASE (GS5)* as a positive regulator of rice grain size (Li Y. et al., 2011), but also an ortholog of *SERINE CARBOXYPEPTIDASE-LIKE 40 (GRMZM2G072240)* involving in formation of the number of hull layers in maize (Cui et al., 2018). Several QTLs for panicle shape, thousand grain weight,

and grain yield were consistently mapped in the same genomic regions as in the previous studies in sorghum (Mantilla Perez et al., 2014; Zhou et al., 2019). Moreover, many QTLs related to disease resistance and morphology were also reported in this genomic region, such as shoot fly resistance (Aruna et al., 2011), ergot resistance (Parh et al., 2008), downy mildew resistance (Ahn et al., 2019), rust resistance (Wang et al., 2014), plant height (Shiringani et al., 2010; Mantilla Perez et al., 2014), tillering number (Alam et al., 2014; Kong et al., 2015), polyphenol content (Rhodes et al., 2017). This is in line with the main traits of Chishui sorghum varieties, which possess smaller grain size, looser panicle, better disease resistance and higher plant height, to adapt to local agroecological environments, such as hot and humid summer, serious pests and diseases. Compared with the thickness of pericarp, it is easier for local farmers to select new varieties based on traits such as grain size, panicle shape, and disease resistance during the domestication and improvement of sorghum. As a result, the strong artificial selection of grain size, panicle type or resistance to biotic stress is most likely to cause the selective sweep, which led to a genetic hitchhiking of nearby *Z* gene conferring pericarp thickness and a large reduction of the genetic diversity in the 140 kb regions on chromosome 2.

## Relationship Between Baijiu-Brewing Process and the Physical Properties of Grains

Traditionally, Moutai-aroma liquor brewing has always preferred glutinous sorghum varieties with smaller size and thicker pericarp, which were thought to meet the process technology of repeated steaming (Chen and Ji, 2006). Contrary to the conventional belief, our results revealed that the increase in pericarp thickness results in larger grain size, and that pericarp thickness rather than grain size is associated with the rate of expansion and crack in waxy sorghum, meaning that smaller-size grains have thinner pericarp, while grain with thinner pericarp generally exhibit better resistance to cooking. The similar results were reported in the effect of grain physical properties on sorghum processing. A thick pericarp absorbs more water and is more easily detached during the dehulling than a thin pericarp (Scheuring et al., 1983). Our findings also suggest that selection for *Z* gene should be avoided in sorghum high-yield breeding program, because the high-yielding varieties caused by thick pericarp have lower rate of dehulled grains and flour yield during milling (Lu, 1999).

In summary, population structure analysis revealed that 333 sorghum accessions clustered into six groups of wild relatives, cultivated sorghum in Africa, Southern Asia, Southern China, Northern China, and Chishui in Southwest China, supporting the hypothesis that sorghum entered China through the Southwest Silk Road. We discovered a significant selective sweep that harbors several genes for pericarp thickness, grain size, panicle

type or biotic stress tolerance, which resulted in a large reduction of the genetic diversity in the 140 kb regions on chromosome 2. Based on the cooking experiment, we put forward a more scientific standard for sorghum breeding that is suitable for liquor brewing technology in China. Our results provided useful information for the breeding and improvement not only for liquor-sorghum but also for other breeding programs in China.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The name of the repository and accession number can be found below: China National GeneBank (CNGB); CNP0002968.

## AUTHOR CONTRIBUTIONS

LZ and GZ designed and managed the research and wrote the manuscript. YD, JX, XG, KL, ZF, NC, BC, LBZ, MR, and ZX collected the plant materials, investigated traits, and performed the experiments. YD, XL, and ZB analyzed whole-genome sequence data and RNA-seq data. ZX and YT revised the manuscript. All authors read and approved the final manuscript.

## FUNDING

This work was supported by the National Natural Science Foundation of China (32160459), the National Key R&D Program of China (2018YFD1000706/2018YFD10007011), Guizhou Natural Science Foundation of China [QKHJC(2020)1Y103], Zhejiang Major Scientific and Technological Project of Agricultural (Upland crop) Breeding (2021C02064-6).

## ACKNOWLEDGMENTS

We thank Liu Shuo from Liaoning Institute of Pomology, China, for the enthusiasm and support in bio-information analysis and revision on the manuscript. We also thank Li Guangwei from Henan Agricultural University, China, for the comments on the manuscript.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.923734/full#supplementary-material>

## REFERENCES

- Adewale, S., Badu-Apraku, B., Akinwale, R., Paterne, A., Gedil, M., and Garcia-Oliveira, A. (2020). Genome-wide association study of *Striga* resistance in early maturing white tropical maize inbred lines. *BMC Plant Biol.* 20:203. doi: 10.1186/s12870-020-02360-0
- Ahn, E., Hu, Z., Perumal, R., Prom, L., Odvody, G., Upadhyaya, H., et al. (2019). Genome wide association analysis of sorghum mini core lines regarding

- anthracnose, downy mildew, and head smut. *PLoS One* 14:e0216671. doi: 10.1371/journal.pone.0216671
- Alam, M., Mace, E., van Oosterom, E., Cruickshank, A., Hunt, C., Hammer, G., et al. (2014). QTL analysis in multiple sorghum populations facilitates the dissection of the genetic and physiological control of tillering. *Theoret. Appl. Genet.* 127, 2253–2266. doi: 10.1007/s00122-014-2377-9
- Aruna, C., Bhagwat, V., Madhusudhana, R., Sharma, V., Hussain, T., Ghorade, R., et al. (2011). Identification and validation of genomic regions that affect shoot fly resistance in sorghum [*Sorghum bicolor* (L.) Moench]. *Theoret. Appl. Genet.* 122, 1617–1630. doi: 10.1007/s00122-011-1559-y
- Awika, J., and Rooney, L. (2004). Sorghum phytochemicals and their potential impact on human health. *Phytochemistry* 65, 1199–1221. doi: 10.1016/j.phytochem.2004.04.001
- Aya, K., Hobo, T., Sato-Izawa, K., Ueguchi-Tanaka, M., Kitano, H., and Matsuoka, M. (2014). A novel AP2-type transcription factor, SMALL ORGAN SIZE1, controls organ size downstream of an auxin signaling pathway. *Plant Cell Physiol.* 55, 897–912. doi: 10.1093/pcp/pcu023
- Ayyangar, G., and Ayyar, M. (1938). Linkage between a panicle factor and the pearly-chalky mesocarp factor (Zz) in sorghum. *Proc. Ind. Acad. Sci. Sect. B.* 8, 100–107. doi: 10.1007/BF03048478
- Boivin, K., Deu, M., Rami, J. F., Trouche, G., and Hamon, P. (1999). Towards a saturated sorghum map using RFLP and AFLP markers. *Theoret. Appl. Genet.* 98, 320–328. doi: 10.1007/s001220051076
- Boivin, N., and Fuller, D. Q. (2009). Shell middens, ships and seeds: exploring coastal subsistence, maritime trade and the dispersal of domesticates in and around the ancient Arabian Peninsula. *J. World Prehist.* 22, 113–180. doi: 10.1007/s10963-009-9018-2
- Chen, B. R., Wang, C. Y., Wang, G. P., Zhu, Z. X., Xu, N., Shi, G. S., et al. (2019). Genome-wide association study for starch content and constitution in sorghum (*Sorghum bicolor* (L.) Moench). *J. Integr. Agricult.* 18, 2446–2456. doi: 10.1016/S2095-3119(19)62631-6
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Chen, X. X., and Ji, K. L. (2006). General Introduction to the individualities of Maotai Liquor. *Liquor-Mak. Sci. Technol.* 140:5.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Cui, Z., Xia, A., Zhang, A., Luo, J., Yang, X., Zhang, L., et al. (2018). Linkage mapping combined with association analysis reveals QTL and candidate genes for three husk traits in maize. *Theoret. Appl. Genet.* 131, 2131–2144. doi: 10.1007/s00122-018-3142-2
- Dahlberg, J. A., and Wasylkowska, K. (1996). Image and statistical analyses of early sorghum remains (8000 b.p.) from the Nabta Playa archaeological site in the Western Desert, southern Egypt. *Veget. Hist. Archaeobot.* 5:6.
- Danecek, P., Auton, A., Abecasis, G., Albers, C., Banks, E., DePristo, M., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330
- Doggett, H. (1988). Sorghum. *Econ. Bot.* 1, 355–371.
- Dong, S., He, W., Ji, J., Zhang, C., Guo, Y., and Yang, T. (2021). LDBlockShow: a fast and convenient tool for visualizing linkage disequilibrium and haplotype blocks based on variant call format files. *Brief. Bioinform.* 22:227. doi: 10.1093/bib/bbaa227
- Dossa, K., Zhou, R., Li, D., Liu, A., Qin, L., Mmadi, M. A., et al. (2021). A novel motif in the 5'-UTR of an orphan gene 'Big Root Biomass' modulates root biomass in sesame. *Plant Biotechnol. J.* 19, 1065–1079. doi: 10.1111/pbi.13531
- Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., et al. (2018). Resequencing of 243 diploid cotton accessions based on an updated A genome identifies the genetic basis of key agronomic traits. *Nature genetics* 50, 796–802. doi: 10.1038/s41588-018-0116-x
- Dufour, P., Deu, M., Grivet, L., D'Hont, A., Paulet, F., Bouet, A., et al. (1997). Construction of a composite sorghum genome map and comparison with sugarcane, a related complex polyploid. *Theoret. Appl. Genet.* 94, 409–418. doi: 10.1007/s001220050430
- Fan, W., Lu, J., Pan, C., Tan, M., Lin, Q., Liu, W., et al. (2019). Sequencing of Chinese castor lines reveals genetic signatures of selection and yield-associated loci. *Nat. Comm.* 10:3418. doi: 10.1038/s41467-019-11228-3
- Fuller, D. Q. (2003). African crops in prehistoric South Asia: A critical review. In: Neumann, K., Butler, A., Kahlheber, S. (Eds.), *Food, Fuel and Fields: Progress in Africa Archaeobotany. Africa Praehist.* 15:32.
- Goodstein, D., Shu, S., Howson, R., Neupane, R., Hayes, R., Fazo, J., et al. (2012). Phytosome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40, D1178–D1186. doi: 10.1093/nar/gkr944
- Groppi, A., Liu, S., Cornille, A., Decroocq, S., Bui, Q., Tricon, D., et al. (2021). Population genomics of apricots unravels domestication history and adaptive events. *Nat. Comm.* 12:3956. doi: 10.1038/s41467-021-24283-6
- Han, M. L. (2012). *Historical Agricultural Geography of China*. Beijing: Peking University Press.
- Hao, C., Jiao, C., Hou, J., Li, T., Liu, H., Wang, Y., et al. (2020). Resequencing of 145 Landmark Cultivars Reveals Asymmetric Sub-genome Selection and Strong Founder Genotype Effects on Wheat Breeding in China. *Mole. Plant* 13, 1733–1751. doi: 10.1016/j.molp.2020.09.001
- Hausmann, B., Hess, D., Omany, G., Folkertsma, R., Reddy, B., Kayentao, M., et al. (2004). Genomic regions influencing resistance to the parasitic weed *Striga hermonthica* in two recombinant inbred populations of sorghum. *Theoret. Appl. Genet.* 109, 1005–1016. doi: 10.1007/s00122-004-1706-9
- Hinnebusch, A., Ivanov, I., and Sonenberg, N. (2016). Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* 352, 1413–1416. doi: 10.1126/science.aad9868
- Hufford, M., Xu, X., van Heerwaarden, J., Pyhäjärvi, T., Chia, J., Cartwright, R., et al. (2012). Comparative population genomics of maize domestication and improvement. *Nat. Genet.* 44, 808–811. doi: 10.1038/ng.2309
- Joseph, N. (2000). *Science and civilisation in China*. Cambridge, MA: Cambridge University Press.
- Kong, W., Kim, C., Goff, V., Zhang, D., and Paterson, A. (2015). Genetic analysis of rhizomatousness and its relationship with vegetative branching of recombinant inbred lines of *Sorghum bicolor* × *S. propinquum*. *Am. J. Bot.* 102, 718–724. doi: 10.3732/ajb.1500035
- Letunic, I., and Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324
- Li, M., Bahn, S., Guo, L., Musgrave, W., Berg, H., Welti, R., et al. (2011). Patatin-related phospholipase pPLAIIIβ-induced changes in lipid metabolism alter cellulose content and cell elongation in Arabidopsis. *Plant Cell* 23, 1107–1123. doi: 10.1105/tpc.110.081240
- Li, R. Y., Zhang, H., Zhou, X. C., Guan, Y. N., Yao, F. X., Song, G. A., et al. (2010). Genetic diversity in Chinese sorghum landraces revealed by chloroplast simple sequence repeats. *Genet. Resour. Crop Evol.* 57, 1–15. doi: 10.1007/s10722-009-9446-y
- Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural variation in *G55* plays an important role in regulating grain size and yield in rice. *Nat. Genet.* 43, 1266–1269. doi: 10.1038/ng.977
- Liu, M. K., Tang, Y. M., Ren, D. Q., Yao, W. C., Ni, B., and Lei, G. D. (2012). Brewing property comparison between several species of sorghum seed. *China Brew.* 31:4.
- Lu, Q. S. (1999). *Sorghum*. Beijing: China Agriculture Press.
- Mace, E., Innes, D., Hunt, C., Wang, X., Tao, Y., Baxter, J., et al. (2019). The Sorghum QTL Atlas: a powerful tool for trait dissection, comparative genomics and crop improvement. *Theoret. Appl. Genet.* 132, 751–766. doi: 10.1007/s00122-018-3212-5
- Mace, E. S., Tai, S., Gilding, E. K., Li, Y., Prentiss, P. J., Bian, L., et al. (2013). Whole-genome sequencing reveals untapped genetic potential in Africa's indigenous cereal crop sorghum. *Nat. Comm.* 4:2320. doi: 10.1038/ncomms3320
- Mann, J. A., Kimber, C. T., and Miller, F. R. (1983). The origin and early cultivation of sorghums in Africa. *Bulletin - Texas Agricult. Exp. Stat.* 1983:21. doi: 10.1007/s10437-018-9314-2
- Mantilla Perez, M., Zhao, J., Yin, Y., Hu, J., and Salas Fernandez, M. (2014). Association mapping of brassinosteroid candidate genes and plant architecture in a diverse panel of Sorghum bicolor. *Theoret. Appl. Genet.* 127, 2645–2662.
- McCormick, R., Truong, S., Sreedasyam, A., Jenkins, J., Shu, S., Sims, D., et al. (2018). The Sorghum bicolor reference genome: improved assembly, gene annotations, a transcriptome atlas, and signatures of genome organization. *Plant J.* 93, 338–354. doi: 10.1111/tpj.13781

- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303. doi: 10.1101/gr.107524.110
- Morris, G. P., Ramu, P., Deshpande, S. P., Hash, C. T., Shah, T., Upadhyaya, H. D., et al. (2013). Population genomic and genome-wide association studies of agroclimatic traits in sorghum. *Proc. Natl. Acad. Sci.* 110, 453–458. doi: 10.1073/pnas.1215985110
- Parh, D., Jordan, D., Aitken, E., Mace, E., Jun-ai, P., McIntyre, C., et al. (2008). QTL analysis of ergot resistance in sorghum. *Theoret. Appl. Genet.* 117, 369–382. doi: 10.1007/s00122-008-0781-8
- Paterson, A. H., Bowers, J. E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., et al. (2009). The Sorghum bicolor genome and the diversification of grasses. *Nature* 457, 551–556. doi: 10.1038/nature07723
- Price, A., Patterson, N., Plenge, R., Weinblatt, M., Shadick, N., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* 38, 904–909. doi: 10.1038/ng1847
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Qiao, Y., Piao, R., Shi, J., Lee, S., Jiang, W., Kim, B., et al. (2011). Fine mapping and candidate gene analysis of dense and erect panicle 3, *DEP3*, which confers high grain yield in rice (*Oryza sativa* L.). *Theoret. Appl. Genet.* 122, 1439–1449. doi: 10.1007/s00122-011-1543-6
- Raj, A., Stephens, M., and Pritchard, J. (2014). fastSTRUCTURE: variational inference of population structure in large SNP data sets. *Genetics* 197, 573–589. doi: 10.1534/genetics.114.164350
- Rami, J. F., Dufour, P., Trouche, G., Fliedler, G., Mestres, C., Davrieux, F., et al. (1998). Quantitative trait loci for grain quality, productivity, morphological and agronomical traits in sorghum (*Sorghum bicolor* L. Moench). *Theoret. Appl. Genet.* 97, 605–616. doi: 10.1007/s001220050936
- Rhodes, D., Gadgil, P., Perumal, R., Tesso, T., and Herald, T. J. (2017). Natural Variation and Genome-Wide Association Study of Antioxidants in a Diverse Sorghum Collection. *Cereal Chem.* 94, 190–198. doi: 10.1094/CCHEM-03-16-0075-R
- Scheuring, J. F., Sidibe, S., Rooney, L. W., and Earp, C. F. (1983). Sorghum pericarp thickness and its relation to decortication in a wooden mortar and pestle. *Cer. Chem.* 60, 86–89.
- Shi, C., Dong, N., Guo, T., Ye, W., Shan, J., and Lin, H. (2020). A quantitative trait locus GW6 controls rice grain size and yield through the gibberellin pathway. *Plant J.* 103, 1174–1188. doi: 10.1111/tpj.14793
- Shinoda. (1958). *Baijiu – The introduction of sorghum Taiwan*. Elite Publishing Co., Ltd.
- Shiringani, A., Frisch, M., and Friedt, W. (2010). Genetic mapping of QTLs for sugar-related traits in a RIL population of *Sorghum bicolor* L. Moench. *Theoret. Appl. Genet.* 121, 323–336. doi: 10.1007/s00122-010-1312-y
- Smith, O., Nicholson, W., Kistler, L., Mace, E., Clapham, A., Rose, P., et al. (2019). A domestication history of dynamic adaptation and genomic deterioration in Sorghum. *Nat. Plants* 5, 369–379. doi: 10.1038/s41477-019-0397-9
- Sui, Z., Wang, T., Li, H., Zhang, M., Li, Y., Xu, R., et al. (2016). Overexpression of Peptide-Encoding *OsCEP6.1* Results in Pleiotropic Effects on Growth in Rice (*O. sativa*). *Front. Plant Sci.* 7:228. doi: 10.3389/fpls.2016.00228
- Sun, B. G. (2019). *Chinese National Alcohols*. Beijing: China Chemical Industry Press.
- Tao, Y., Trusov, Y., Zhao, X., Wang, X., Cruickshank, A., Hunt, C., et al. (2021). Manipulating assimilate availability provides insight into the genes controlling grain size in sorghum. *Plant J.* 108, 231–243. doi: 10.1111/tpj.15437
- Tao, Y. Z., Jordan, D. R., Hanzell, R. G., and McIntyre, C. L. (1998). Construction of a genetic map in a sorghum recombinant inbred line using probes from different sources and its comparison with other sorghum maps. *Austr. J. Agric. Res.* 49, 729–736. doi: 10.1071/A97112
- Varshney, R. K., Thudi, M., Roorkiwal, M., He, W., Upadhyaya, H. D., Yang, W., et al. (2019). Resequencing of 429 chickpea accessions from 45 countries provides insights into genome diversity, domestication and agronomic traits. *Nat. Genet.* 51, 857–864. doi: 10.1038/s41588-019-0401-3
- Wang, X., Mace, E., Hunt, C., Cruickshank, A., Hanzell, R., Parkes, H., et al. (2014). Two distinct classes of QTL determine rust resistance in sorghum. *BMC Plant Biol.* 14:366. doi: 10.1186/s12870-014-0366-4
- Winchell, F., Stevens, C. J., Murphy, C., Champion, L., and Fuller, D. Q. (2017). Evidence for Sorghum Domestication in Fourth Millennium BC Eastern Sudan Spikelet Morphology from Ceramic Impressions of the Butana Group. *Curr. Anthropol.* 58, 673–683.
- Wu, J., Wang, Y., Xu, J., Korban, S. S., Fei, Z., Tao, S., et al. (2018). Diversification and independent domestication of Asian and European pears. *Gen. Biol.* 19:77. doi: 10.1186/s13059-018-1452-y
- Xu, W., Wu, D., Yang, T., Sun, C., Wang, Z., Han, B., et al. (2021). Genomic insights into the origin, domestication and genetic basis of agronomic traits of castor bean. *Gen. Biol.* 22:113. doi: 10.1186/s13059-021-02333-y
- Xu, X., Liu, X., Ge, S., Jensen, J., Hu, F., Li, X., et al. (2011). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomically important genes. *Nat. Biotechnol.* 30, 105–111. doi: 10.1038/nbt.2050
- Ye, J., Li, W., Ai, G., Li, C., Liu, G., Chen, W., et al. (2019). Genome-wide association analysis identifies a natural variation in basic helix-loop-helix transcription factor regulating ascorbate biosynthesis via D-mannose/L-galactose pathway in tomato. *PLoS Genet.* 2019:15. doi: 10.1371/journal.pgen.1008149
- You, X. L. (1989). The time and origin of the introduction of corn into China and Asia. *Ancient Mod. Agricult.* 2:9.
- Zeng, X., Guo, Y., Xu, Q., Mascher, M., Guo, G., Li, S., et al. (2018). Origin and evolution of qingke barley in Tibet. *Nat. Comm.* 9:5433. doi: 10.1038/s41467-018-07920-5
- Zhang, C., Dong, S., Xu, J., He, W., and Yang, T. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875
- Zhang, H., Wang, J.-C., Wang, D.-J., Yao, F.-X., Xu, J.-F., Song, G.-A., et al. (2011). Assessment of genetic diversity in Chinese sorghum landraces using SSR markers as compared with foreign accessions. *Crop J.* 37, 224–234. doi: 10.3724/SP.J.1006.2011.00224
- Zhang, L., Leng, C., Luo, H., Wu, X., Liu, Z., Zhang, Y., et al. (2018). DrySweet Sorghum Originated through Selection of a Plant-Specific NAC Transcription Factor Gene. *Plant Cell* 30, 2286–2307. doi: 10.1105/tpc.18.00313
- Zhao, L. J. (2019b). Study on Ancient Names of Sorghum in the Pre-Qin and Han Dynasties. *Agricult. Archaeol.* 2019, 36–45.
- Zhao, L. J. (2019a). Discuss on the Time, Path of Sorghum's Introduction into China and its Preliminary Popularization. *Agricult. Hist. Chin.* 2019:1.
- Zhao, Y. P., Fan, G., Yin, P. P., Sun, S., Li, N., Hong, X., et al. (2019). Resequencing 545 ginkgo genomes across the world reveals the evolutionary history of the living fossil. *Nat. Comm.* 10:4201. doi: 10.1038/s41467-019-12133-5
- Zhou, Y., Srinivasan, S., Mirnezami, S., Kusmec, A., Fu, Q., Attigala, L., et al. (2019). Semiautomated Feature Extraction from RGB Images for Sorghum Panicle Architecture GWAS. *Plant Physiol.* 179, 24–37. doi: 10.1104/pp.18.00974
- Zou, G., Zhai, G., Yan, S., Li, S., Zhou, L., Ding, Y., et al. (2020). Sorghum qTGW1a encodes a G-protein subunit and acts as a negative regulator of grain size. *J. Exp. Bot.* 71, 5389–5401. doi: 10.1093/jxb/eraa277

**Conflict of Interest:** ZB was employed by Shanghai OE Biotech Co., Ltd., China.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Zhang, Ding, Xu, Gao, Cao, Li, Feng, Cheng, Zhou, Ren, Lu, Bao, Tao, Xin and Zou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.