Molecular Plant
Research article

CellPress
Partner Journal

Genomic footprints of Kam Sweet Rice domestication indicate possible migration routes of the Dong people in China and provide resources for future rice breeding

Chunhui Liu[1,2,5], Tianyi Wang[3,5], Huicha Chen[4,5], Xiaoding Ma[1], Chengzhi Jiao[1], Di Cui[1], Bing Han[1], Xiaobing Li[4], Aixia Jiao[4], Renchao Ruan[4], Dayuan Xue[2], Yanjie Wang[1,*] and Longzhi Han[1,*]

[1]Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China

[2]College of Life and Environmental Sciences, Minzu University of China, Beijing 100081, China

[3]Smartgenomics Technology Institute, Tianjin 301700, China

[4]Institute of Crop Germplasm Resources, Guizhou Academy of Agricultural Sciences, Guiyang 550025, China

[5]These authors contributed equally to this article.

***Correspondence: Yanjie Wang (wangyanjie@caas.cn), Longzhi Han (hanlongzhi@caas.cn)**

https://doi.org/10.1016/j.molp.2022.12.020

ABSTRACT

The Dong people are one of China's 55 recognized ethnic minorities, but there has been a long-standing debate about their origins. In this study, we performed whole-genome resequencing of Kam Sweet Rice (KSR), a valuable, rare, and ancient rice landrace unique to the Dong people. Through comparative genomic analyses of KSR and other rice landraces from south of the Yangtze River Basin in China, we provide evidence that the ancestors of the Dong people likely originated from the southeast coast of China at least 1000 years ago. Alien introgression and admixture in KSR demonstrated multiple migration events in the history of the Dong people. Genomic footprints of domestication demonstrated characteristics of KSR that arose from artificial selection and geographical adaptation by the Dong people. The key genes *GS3*, *Hd1*, and *DPS1* (related to agronomic traits) and *LTG1* and *MYBS3* (related to cold tolerance) were identified as domestication targets, reflecting crop improvement and changes in the geographical environment of the Dong people during migration. A genome-wide association study revealed a candidate yield-associated gene, *Os01g0923300*, a specific haplotype in KSR that is important for regulating grain number per panicle. RNA-sequencing and quantitative reverse transcription–PCR results showed that this gene was more highly expressed in KSR than in ancestral populations, indicating that it may have great value in increasing yield potential in other rice accessions. In summary, our work develops a novel approach for studying human civilization and migration patterns and provides valuable genomic datasets and resources for future breeding of high-yield and climate-resilient rice varieties.

Key words: Dong ethnic group, Kam Sweet Rice, genomic evolution, population genetics

Liu C., Wang T., Chen H., Ma X., Jiao C., Cui D., Han B., Li X., Jiao A., Ruan R., Xue D., Wang Y., and Han L. (2023). Genomic footprints of Kam Sweet Rice domestication indicate possible migration routes of the Dong people in China and provide resources for future rice breeding. Mol. Plant. **16**, 415–431.

INTRODUCTION

Human migration has changed the world as well as humanity itself (Kutanan et al., 2021; Ning et al., 2020). All extant ethnic groups in China migrated continuously throughout history and eventually formed concentrated communities (Oliveira et al., 2022). The Dong people, one of China's 55 ethnic minorities, are a large group, with a total population of 3.5 million. Southeast Guizhou, also known as the Qiandongnan area (QDN), has the largest population of Dong people in China. However, there has been

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and CEMPS, CAS.

extensive debate among historians, archaeologists, and anthropologists about the origin and migration of the Dong people. Four main possible migration routes have been proposed (Supplemental Figure 1A and 1B and Supplemental Note). All possible proposed origins have something in common: they all belong to regions that harbored Yangtze River or Pearl River civilizations from the Neolithic period. These civilizations are famous for their rich cultures, which relied on rice cultivation (Larson et al., 2014) (Supplemental Note). Archaeological evidence shows that the Dong people are one of the oldest rice-farming ethnic groups in China (You and Zeng, 2010). They mainly planted rice resources referred to as "HE" in the local dialect for thousands of years (Wang et al., 2018a; Li et al., 2019). Because the Dong people call themselves the Kam or Kam people, HE resources are collectively referred to as Kam Sweet Rice (KSR) (Supplemental Figure 2), which is the international standard translation reported by *Science* (Stone, 2008). It is also called "Specialty Rice" by the Food and Agriculture Organization of the United Nations (Bedigian, 2003). In China, scholars defined KSR as a traditional landrace formed through long-term natural and artificial selection in a complex and diverse ecological environment as a result of the traditional farming methods of the Dong people (Ma, 1979).

To use KSR genetic evolution data to infer migration patterns of the Dong people, three conditions must be met in the relationship between the plant and the people, which are met in this case. First, KSR is cultivated and traditionally used only by the Dong people. Historical and linguistic data confirm that the pronunciation of HE in the QDN dialect comes from the Dong group of the Sino-Tibetan phylum of languages (Yang, 2013). This means that the Dong people are the most primitive cultivators of KSR (Wang et al., 2018a) (Supplemental Note). When social exchanges flourish, it is common for different ethnic groups to live in the same area. However, our ethnological investigation found that KSR was cultivated only by the Dong people (Liu et al., 2022a) (Condition 1). Second, KSR was an essential material carried by the Dong people during their migrations and then domesticated in the QDN area to fit the local environment (Yang, 2021). Anthropologists have demonstrated a historical center of glutinous rice culture in South China (Huang, 2016). The ancestors of the Dong people who lived in this center later brought rice farming to new locations as they migrated (Yang, 2014) (Condition 2), and traditional Dong culture emerged with KSR at the core (Pan and Long, 2013). Third, the Dong people did not use any local rice landraces in Guizhou or other nearby areas (Lei et al., 2021; Liu et al., 2022a). They live in remote mountain valleys along meandering streams in Southwest China, where there was no transportation a decade ago; the Dong people were thus virtually isolated, and very few other rice landraces have been introduced to QDN. This means that only KSR has been planted by the Dong people (Ruan et al., 2007) (Condition 3). Based on these conditions, it is feasible to use KSR genomics to accurately hypothesize about the migratory activities of the Dong people.

Domestication has profound effects on crop genomes (Cui et al., 2022; Wu et al., 2022). The genomic footprints of KSR domestication and breeding as a result of selection by the Dong people have been preserved. KSR is suitable for growth in cold areas such as mountains and valleys and without pesticides or fertilizers (Wang et al., 2018a; Liu et al., 2022a). Thus, KSR has strong resistance to cold, disease, pests, and stress (Liu et al., 2022b). KSR is also notable for yield traits such as large grain size and high seed-setting rates. The Dong people have carefully chosen varieties with round, plump grains and high seed-setting rates as the main planting varieties based on their traditional cultural customs and livelihood needs (Supplemental Note). In addition, landraces with early heading dates were continuously selected by the Dong people during domestication; they believed that early-maturing varieties were crucial for enhancing yield stability by avoiding low temperatures. Our ethnobiological study found that the Dong people have a unique set of traditional technical methods for KSR domestication and breeding selection, as detailed in the Supplemental Note. However, the mechanisms by which these processes have shaped KSR genomes are not fully understood, as discussed in the present study.

We previously conducted a simple-sequence-repeat marker study in KSR and other rice landraces from Guizhou and surrounding provinces. Surprisingly, the genetic distance was closer between KSR and landraces from Guangdong and Jiangxi provinces than between KSR and other landraces from Guizhou province (Liu et al., 2022b), and this is no coincidence. This finding indicated that there may be an evolutionary relationship between KSR and rice landraces from surrounding provinces, which we attempted to use to seek clues about Dong migration, thus providing molecular biological evidence for the migration of the Dong people and also a reference for the study of human origins and migration. Round and full grains, high seed-setting rate, and short growth period are the unique characteristics that distinguish KSR from other rice landraces (Wu et al., 2010; Li et al., 2019). Clearly, KSR is a very precious material for rice breeding. Unfortunately, the planting area and number of KSR varieties have sharply decreased in recent years, and KSR is now threatened with extinction (Lei et al., 2017). Exploring the genetic evolution of KSR is an essential way to protect this unique collection of excellent rice landraces.

Here, a total of 439 accessions were resequenced (Supplemental Figure 3). Group 1 comprised 104 KSR accessions from QDN. Group 2 contained 104 Guizhou (GZ) landraces from areas in Guizhou Province outside of QDN. Group 3 comprised 164 landraces from south of the Yangtze River, including accessions from Central China (CC), East China (EC), South China (SC), and Southwest China (SW) (Supplemental Figure 1C). Group 4 contained 37 wild rice varieties from countries where *Oryza rufipogon* was widely distributed. Group 5 contained 15 Aus-type varieties from Bangladesh and 15 Basmati and Sadri aromatic varieties from India, Nepal, Bhutan, Liberia, Myanmar, Iran, and Madagascar (Bas). Accessions in Groups 1–3 were resequenced for this study, and genomic data for accessions in Groups 4 and 5 were obtained from previous publications (Wang et al., 2018b; Zhao et al., 2018; Han et al., 2022).

RESULTS

Genome resequencing and single-nucleotide polymorphism identification

To fully determine the origin, domestication history, and genetic basis of KSR, we collected 1481 accessions originating from

south of the Yangtze River in China. This included 1439 landraces (*Oryza sativa*) and 42 wild rice varieties (*O. rufipogon*), all of which were possible genetic donors to KSR. Based on a previous effort to analyze the phenotypes, genetic diversity, and nucleotide variation in these accessions (Liu et al., 2022b), we selected 372 accessions as a core group of morphologically, genetically, and geographically diverse landraces for resequencing (Supplemental Table 1 and Supplemental Figure 4). These 372 accessions represented ~99.5% of the genetic diversity of rice landraces south of the Yangtze River (Supplemental Table 2). Approximately 2.06 Tb of raw sequences with a 150-bp read length were generated, with an average coverage depth of ~12.4× for each accession. We identified 3 566 872 high-quality single-nucleotide polymorphisms (SNPs) (Supplemental Tables 3 and 4).

Population structure analysis suggests that SC and EC accessions are potential donors to KSR

Genetic distances calculated from SNPs were used to determine the phylogenetic relationships of 439 accessions. The *indica* and *japonica* subpopulations were separated, and other subpopulations, including wild rice, Aus, and Bas, were also distinguished in phylogenetic trees (Figure 1A). We identified 163 typical *indica* and 209 typical *japonica* landraces. Most of the *japonica* varieties could be directly connected to their geographic origins (Supplemental Table 1), whereas *indica* varieties could not. This was consistent with results from a previous study on rice landraces in Southwest China (Cui et al., 2021). Although KSR is currently cultivated in Guizhou Province, model-based clustering showed that KSR varieties were grouped together in a single cluster and were significantly different from other rice landraces in Guizhou (Figure 1A). This indicated that long-term natural and artificial selection by the Dong people played an important role in the domestication and spread of KSR (Olsen and Wendel, 2013). KSR varieties are similar to Bas varieties in that both types have a very strong aroma, but they form entirely different subpopulations. To further study the population structure, we used the population clustering program STRUCTURE and performed principal-component analysis (PCA) on the 439 accessions (Supplemental Figure 5). The results from these analyses were similar to those from the phylogenetic tree.

Because ~80.0% of KSR varieties belong to the *japonica* group, we next investigated the relationship between *japonica*-KSR and other *japonica* landraces in greater detail. Genetic diversity was determined by estimating nucleotide diversity (π) (Figure 2A; Supplemental Table 5). The results showed that the nucleotide diversity of KSR ($\pi = 5.77 \times 10^{-4}$) was comparable to that of EC and SC, but lower than that in the other groups. KSR contained approximately 87.01% of the genetic diversity in the GZ group. Compared with the SC and EC accessions, the KSR population had relatively superior agronomic traits (Figure 2C and 2D and Supplemental Figure 6), which were selected by humans in the course of domestication. This result indicated that KSR originated and spread later than EC and SC and that reduced KSR diversity resulted from domestication (Haudry et al., 2007; Du et al., 2018). This finding was consistent with previous results based on molecular diversity using simple sequence repeats (Liu et al., 2022b).

The phylogenetic tree showed that the *japonica* population clustered into six groups: KSR, GZ, CC, EC, SC, and SW (Figure 1B). KSR was derived from Guangdong and Guangxi provinces in the SC group and from Fujian and Jiangxi provinces in the EC group. PCA also showed that EC, SC, and KSR were closely related (Figure 1C). The evolutionary history of KSR was further inferred using individual ancestry coefficients. With K values from 3 to 7, new subpopulations arose from rice landraces (Figure 1D), and PCA confirmed the existence of such subpopulations. When $K = 7$, there was gene exchange between KSR and the SC and EC populations. Furthermore, the population-differentiation statistic (F_{ST}) between KSR and SC was estimated at 0.05, indicating very minor population differentiation (Figure 2B and Supplemental Table 5). By contrast, there was an obvious genetic distinction between KSR and CC and SW, with relatively high population differentiation values (0.14 and 0.20, respectively). Therefore, we preliminarily concluded that the direct donors to KSR were most likely landraces from Guangdong and Guangxi for SC and from Fujian and Jiangxi for EC.

The magnitude of alien introgression in KSR provides evidence for multiple migrations of the Dong people

Results from the above analyses supported Route III (Supplemental Figure 1B) as the migration route of the Dong people. To test this hypothesis, we evaluated gene flow from potential donors to KSR using TreeMix (Pickrell and Pritchard, 2012), the ABBA–BABA test (*D*-statistics) (Durand et al., 2011), and f_d statistics (Martin et al., 2014).

TreeMix was used to examine the topology of relationships and migration history between populations. We observed the oldest split from the common ancestor to be between SW/GZ, SC/KSR, and CC/EC. This was followed by more recent splits between SW and GZ, SC and KSR, and CC and EC (Figure 3A). Support for this model was provided by inferred gene flow from the EC to the KSR and SC groups ($m = 2$) (Figure 3A and Supplemental Figure 7). We then verified gene flow between populations in other provinces and KSR at the genome-wide level using ABBA–BABA statistics; the Z scores from this analysis supported the existence of gene flow ($|Z| > 3$) (Supplemental Table 6). Because the accessions tested were *japonica* landraces from south of the Yangtze River basin, it is reasonable to conclude that there was consistent gene flow between these populations.

To identify specific genomic regions subject to gene flow, f_d statistics were calculated to estimate the magnitude of gene flow (Figure 3B). Of all tested provinces, the population from Guangxi showed the highest f_d values across the genome (*t*-test, $P < 0.05$), followed by Guangdong, Guizhou, Jiangxi, and then Fujian (Figure 3C). We next calculated the proportion of genome introgression to reflect the actual size of the KSR genome derived from alien introgression. The largest proportion of genome introgression identified was from Guangxi (9.14%, 34.11 Mb). The size and frequency of introgressed segments varied across the KSR genome (Supplemental Figure 8). The largest introgressed segment (4.03 Mb) we identified was also from Guangxi; this segment was on chromosome 10 and contained 103 regions with 403 introgressed genes (Supplemental Table 7), covering ~17.4% of the whole

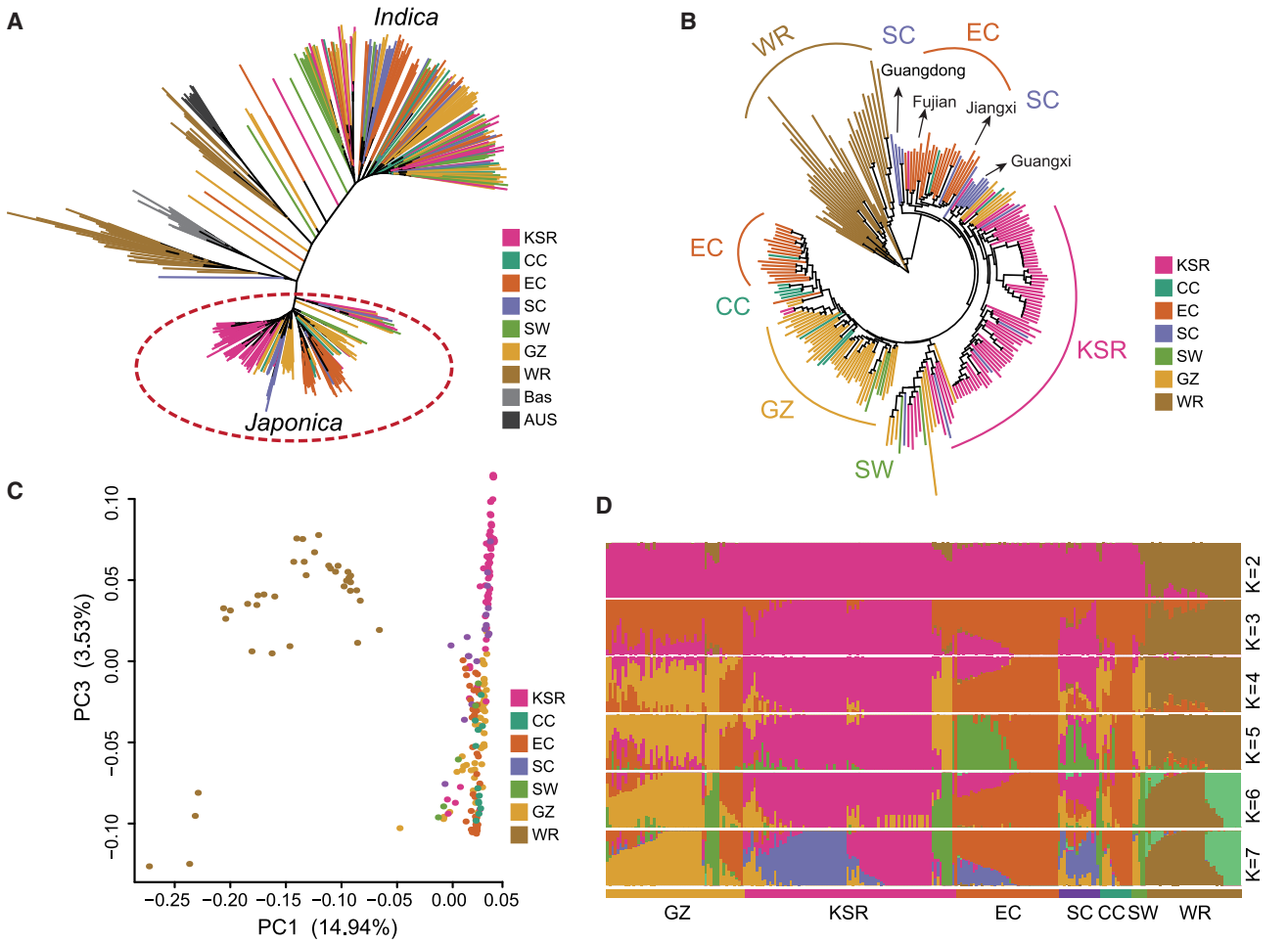

Figure 1. Population structure of rice accessions collected in this study.

(A) Phylogenetic tree of 439 accessions inferred from whole-genome SNPs. The line colors indicate different groups. *Japonica* is encircled with a dashed line; this part of the tree is featured in **(B)**.

(B) The phylogeny of the 209 *japonica* landraces with wild rice as the outgroup.

(C) Principal-component analyses of 209 *japonica* landraces and wild rice.

(D) Individual ancestry coefficients from $K = 2$ to $K = 7$ of 209 *japonica* landraces and wild rice.

chromosome. Jiangxi varieties had the next highest level of introgression (3.90 Mb). This was on chromosome 11 and contained 95 regions with 461 genes. Genomic analysis indicated that the magnitude of alien introgression from Guangxi, Guangdong, Jiangxi, and Fujian provinces was significant and that rice landraces in these provinces were potential donors to KSR, so we inferred the source of the KSR donors from the magnitude of introgression observed (Zhou et al., 2020). These findings were consistent with Routes III and IV, which have been proposed by historians, archaeologists, and anthropologists. Furthermore, these results indicated that the Dong people migrated more than once.

Notably, the ancestral population from which KSR was derived was located on the southeast coast. The characteristic high temperatures and humid heat there are conducive to the occurrence of rice diseases and insects. Some local varieties in these areas are resistant to diseases and insects. As expected, the rice blast resistance genes *Pia* (*Os11g0225100*) and *Pi5* (*Os11g0225300*) and the bacterial blight resistance gene *Xa21* (*Os11g0559200*) (located on chromosome 11) are introgression targets (Figure 3D). Analyses showed that these dominant and rare haplotypes were retained in KSR (Figure 3E).

Population demographic history reveals a genetic bottleneck and separation of KSR

To better understand the demographic history of KSR, we employed SMC++ to infer the effective population size (*Ne*) through time and the divergence times between KSR and the SC and EC populations. The results showed that SC, EC, and WR underwent a significant reduction in *Ne* ~6000–4800 years before present (YBP) (Supplemental Figure 9), suggesting a severe population bottleneck in rice populations. The domestication of cultivated rice was completed by that time, and the rice farming system in China had been established. This has been demonstrated by archaeological data suggesting that rice civilization was formed at around this time in the Yangtze Valley of China (Fuller et al., 2010). After this bottleneck, we observed a gradual increase in *Ne* in the SC and EC varieties, reaching a maximum at

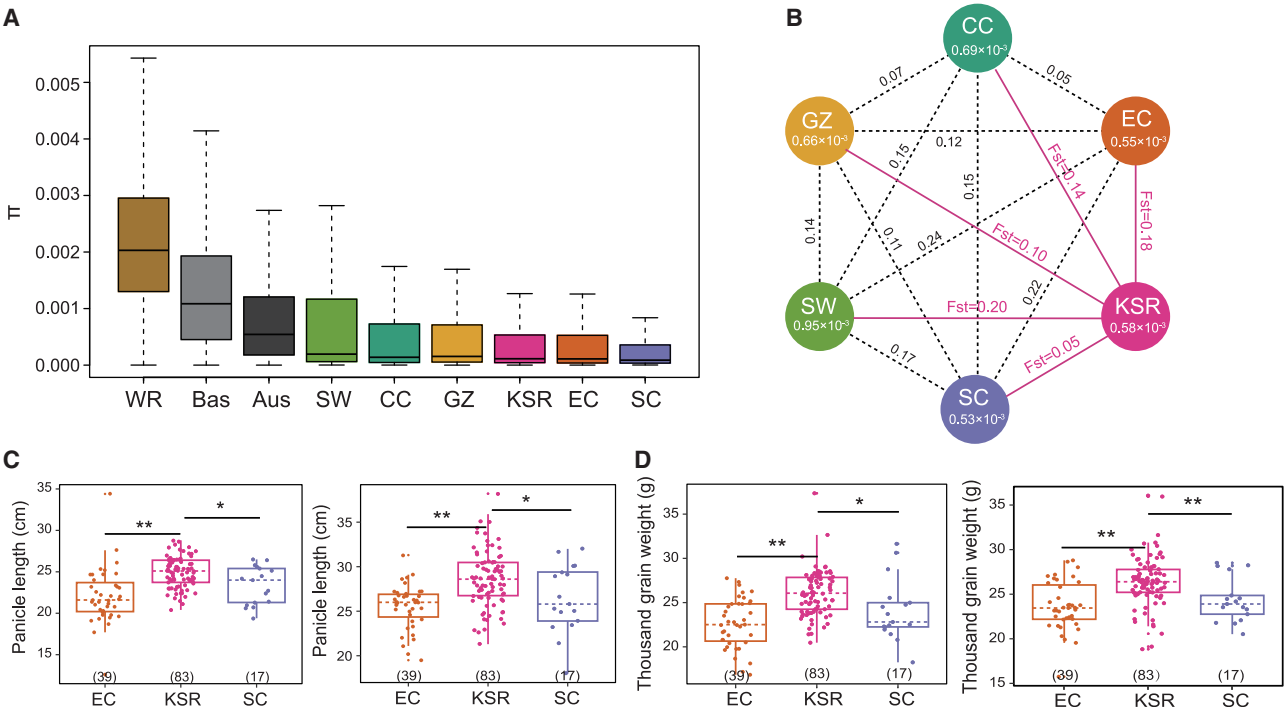

Figure 2. Genetic diversity, population differentiation, and phenotypic distributions.

(A) Nucleotide diversity (π) of *japonica*-KSR, other *japonica* rice landraces, WR, Bas, and Aus.

(B) Genetic diversity and population differentiation between *japonica*-KSR and other *japonica* rice landrace groups. The circles represent the level of genetic diversity (π) of the groups, and the F_{ST} values between the groups are indicated.

(C and D) Phenotypic distributions of KSR, EC, and SC populations. The graphs on the left represent plantings in Sanya, Hainan Province, in 2020; those on the right represent plantings in QDN, Guizhou Province, in 2021. The asterisks indicate significant differences (*$P < 0.05$; **$P < 0.01$).

~180–110 YBP, which could be linked to a population boom in 19th century China (Ge, 2005) (Supplemental Figure 9). The KSR population separated from those of SC and EC ~350 YBP (Figure 3F), which indicated that KSR began to differ from other rice landraces in SC and EC and formed unique population characteristics. Based on ethnological and historical records, the Dong people have cultivated rice in QDN since 1000 YBP, in the Song Dynasty (Yang, 2014), suggesting that domestication of KSR occurred over a long period of ~650 years.

After the separation of KSR, *Ne* gradually increased, reaching a peak at ~220–150 YBP (Figure 3F). This is consistent with historical records showing that the cultivation area and number of KSR varieties reached a peak in the middle of the Qing Dynasty (Yang, 2004). Interestingly, SMC++ results indicated that there was a domestication bottleneck in KSR, with *Ne* starting around 50 000 at ~150 YBP and decreasing to 15 000 at ~120 YBP. The prolonged decrease in *Ne* resulted from a variety of factors. First, from 1870 to 1879 (~140–150 YBP), Guizhou Province suffered from frequent natural disasters, with 57 extreme natural disasters occurring in this period (He, 2011). QDN was the most affected area and suffered from crop failure (Yan, 2009). Second, in the late period of the Qing Dynasty and the early period of the Republic of China (~120 YBP), the government mandated planting of *indica* rice to increase yield (Yang, 2004, 2014; Yan, 2008), which resulted in a significant decrease in the cultivation area of KSR. It is also worth noting that there was a precipitous decline in *Ne* in KSR 30 years ago; this corresponds to the widespread introduction of hybrid rice in the late 1990s to replace landraces (Supplemental Note).

Identification of selective sweeps reveals potential targets of selection during domestication

Genome-wide SNP analysis allowed us to examine not only aspects of the genetic evolution of KSR but also the artificial and natural selection of KSR in the context of traditional Dong culture. We employed three metrics, π values (Pi), a pairwise fixation statistic (F_{ST}), and a cross-population composite likelihood ratio (XP-CLR) to identify putative selective sweeps associated with domestication (Supplemental Table 8). We compared KSR with both the ancestral population (AP), which comprised Guangdong, Guangxi, Jiangxi, and Fujian provinces (AP vs. KSR) (Supplemental Table 9A–9C), and the GZ population (GZ vs. KSR) (Supplemental Table 10A–10C). Due to the potential effects of population structure, bottleneck, or random drift on these results, 651 and 862 genes were considered important candidates for the AP vs. KSR and GZ vs. KSR comparisons, respectively (detected by at least two metrics, these genes were located in the selective sweeps but not in introgression regions) (Supplemental Table 11). These genes had functions related to specialized metabolites, protein processing, and plant–pathogen interactions (Supplemental Figure 10 and Supplemental Table 12).

Domestication of crops by humans focused first on intuitive agronomic traits. Over 2000 plant species have been modified

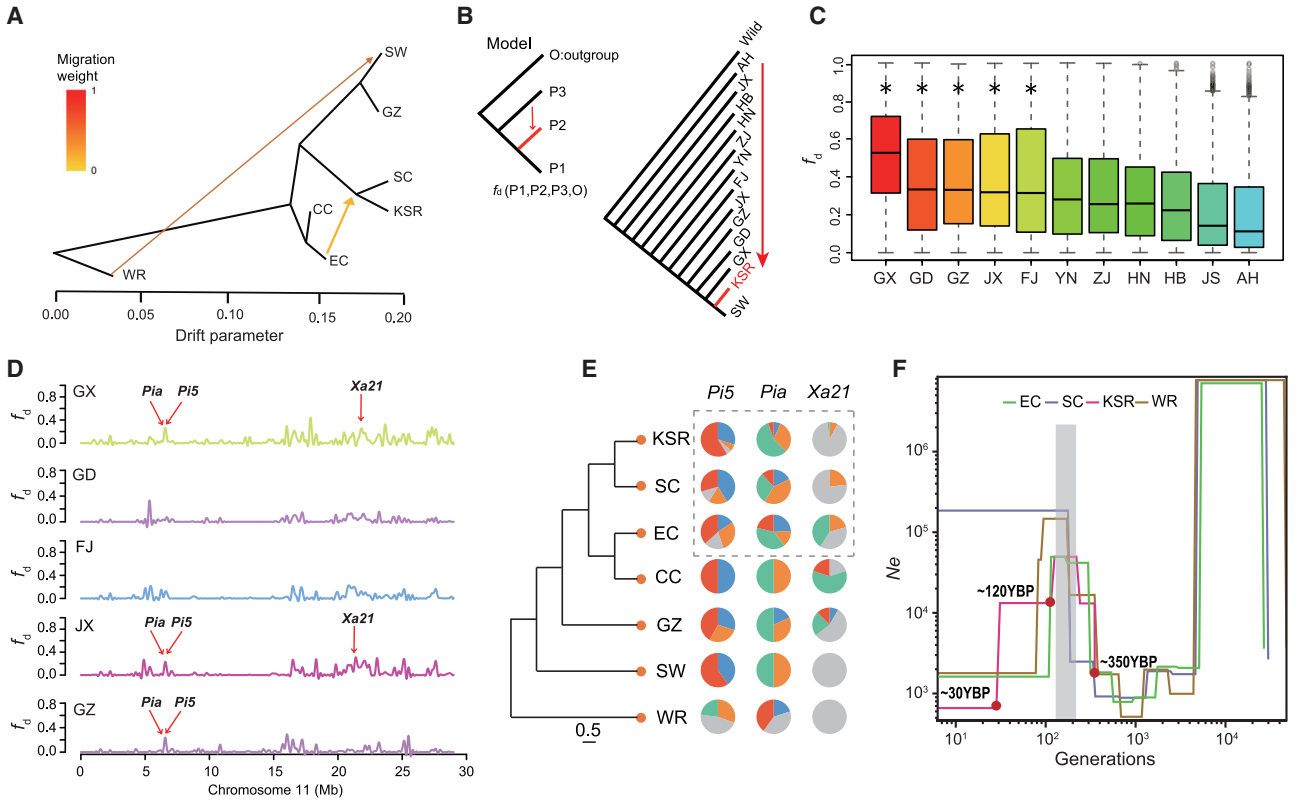

Figure 3. Introgression of *japonica* rice landraces of different groups into KSR.

(A) Maximum likelihood tree inferred by TreeMix allowing two migration events. The arrows correspond to the direction of migration.

(B) The four-taxon topology used for modeling introgression in ABBA–BABA tests (left) and the phylogenetic topology used for inferring introgression to KSR. KSR was used as the recipient population (P2).

(C) Testing gene flow from donor populations to KSR using SW as P1. The asterisk indicates that gene flow (as indicated by f_d) from a specific donor is significantly higher than from Yunnan (YN), Zhejiang (ZJ), Hunan (HN), Hubei (HB), Jiangsu (JS), and Anhui (AH) provinces (*t*-test, $P < 0.05$).

(D) Introgression from donor populations to KSR across chromosome 11. The introgressed genes are labeled at corresponding positions on the chromosome (*Pia*, *Os11g0225100*; *Pi5*, *Os11g0225300*; *Xa21*, *Os11g0559200*). GX, Guangxi; GD, Guangdong; FJ, Fujian; JX, Jiangxi; GZ, Guizhou.

(E) The haplotype analysis of *Pia*, *Pi5*, and *Xa21*. KSR retains dominant and rare haplotypes of ancestral populations (SC and EC).

(F) Effective population size and split time inferred by SMC++ based on WGS SNPs for KSR (pink), EC (green), SC (purple), and WR (brown). The gray bar indicates the maximum effective population size of KSR. YBP, years before present.

morphologically through cultivation and human use (Gaut et al., 2018). As expected, many well-studied genes controlling agronomically important traits were identified in selective sweeps (Figure 4A and Supplemental Figure 11A). For example, in the AP vs. KSR group, genes related to plant architecture and yield were detected. The gene *d50* influences the plant height of rice, and previous studies confirmed that *d50* mutation induced abnormally oriented cell division, resulting in abnormally organized cell files of the internode parenchyma and a dwarf phenotype (Sato-Izawa, et al., 2012). *Ghd7* encodes a transcription factor and is a major gene regulating rice yield, mainly by coordinating plant height, flowering time, and number of grains per spike (Wang et al., 2021; Zong et al., 2021). *GIF1*, which influences grain size and grain filling, has been reported as a potential domestication gene, and such domestication-selected genes can be used for further crop improvement (He et al., 2008). The selection of these genes indicates that key agronomic traits such as plant height, grain size, seed-setting rate, and heading date were selected by the Dong people.

We found that the selected region between 19.23 Mb and 19.26 Mb on chromosome 5 contained the gene *DPS1* (Supplemental Figure 11B), whose superior haplotype predominated in KSR. *DPS1* encodes a mitochondrial-localized protein (containing a cystathionine β-synthase domain) that plays a vital role in the seed-setting rate of rice (Zafar et al., 2020). Haplotype analysis of *DPS1* revealed that GZ, CC, and EC share Hap2, which is absent from SC and KSR (Supplemental Figure 11C), indicating that the KSR and SC populations evolved later than other populations. Phylogenetic trees and haplotype networks showed that Hap1 had the closest genetic evolutionary relationship to Hap3 (Supplemental Figure 11D and 11E), implying that Hap1 was probably domesticated from Hap3. The superior haplotype, Hap1, increased in frequency to become the dominant haplotype (84%) in KSR after domestication (Supplemental Figure 11C and 11F). This indicated that KSR had a high seed-setting rate, which is consistent with previous research (Li et al., 2019). RNA-sequencing (RNA-seq) data showed that *DPS1* had higher expression in KSR than in APs (Supplemental Figure 11G). Next, we used quantitative

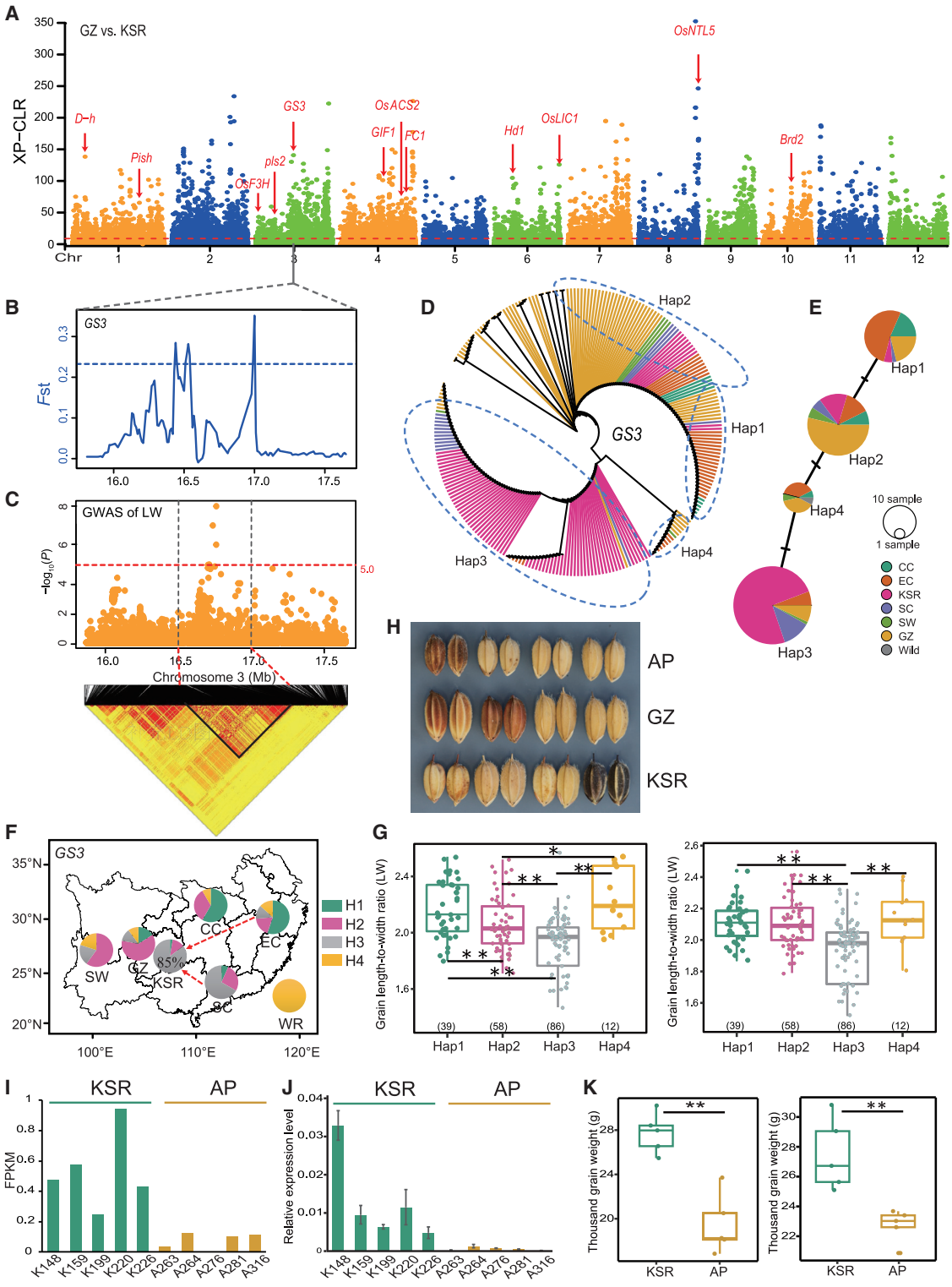

Figure 4. Representative selected gene *GS3*.

(A) Genome-wide selective signals (XP-CLR score) of GZ vs. KSR. Labeled genes were detected by at least two metrics of Pi, F_{ST}, and XP-CLR.

(B) F_{ST} plot for *GS3*.

(C) Local Manhattan plot obtained from GWAS on grain length-to-width ratio (LW) (top) in chromosome 3, with dashed lines indicating the threshold for GWAS ($-\log_{10} P = 5$), and LD heatmap (bottom) surrounding the peak on chromosome 3. Black triangles indicate the positions of the LD block, and the color key indicates r^2 values between SNPs in the regions.

(D) Phylogenetic tree of *GS3*. The line colors indicate different groups.

(legend continued on next page)

real-time PCR to investigate expression levels of *DPS1* and found that it was more abundant in KSR than in APs (Supplemental Figure 11H). Furthermore, KSR has a higher seed-setting rate compared with APs (Supplemental Figure 11I), showing the positive direction of domestication. Thus, we identified *DPS1* as a potential candidate for selection by the Dong people to improve yield during domestication. KSR has a high seed-setting rate because the genome has undergone profound changes due to selection, and this process has left genomic footprints. These changes are closely related to the unique methods of domestication and breeding selection by the Dong people, confirming the results of our ethnobiological investigation at the genomic level. Several other genes associated with agronomic traits were also identified, such as *OsDET1* (Zang et al., 2016), *OsPRA2* (Zhang et al., 2016), *OsNAC20* (Wang et al., 2020), and others. We also identified some receptor-like kinase-related genes such as *OsBRR1* (Peng et al., 2009) and *OsWAK92* (Delteil et al., 2016), which are involved in rice resistance responses to blast fungus. The tyrosine decarboxylase-related gene *OsTyDC1* enhances resistance to rice blast and bacterial blight (Shen et al., 2021). *Bph37* (Zhou et al., 2021), a newly cloned brown planthopper resistance gene on chromosome 6, was also selected. These findings are consistent with the fact that KSR is resistant to diseases and insects when cultivated using primitive cultivation methods, which are not compatible with pesticides and fertilizers (Liu et al., 2022b). Overall, KSR has thus accumulated beneficial mutations that reflect the dietary and economic demands of the Dong people.

In the GZ vs. KSR group, we found that the selected region between 9.33 Mb and 9.34 Mb on chromosome 6 contained a well-known gene, *Hd1*, which influences the heading date of rice plants (Supplemental Figure 12A) (Yano et al., 2000). Previous studies confirmed that *Hd1* has undergone human selection to diversify the flowering times of rice during domestication (Takahashi and Shimamoto, 2011). We observed that the haplotypes of all wild rice were of the undomesticated type, whereas domesticated types were found in cultivated rice. Among them, the superior haplotype, Hap2, evolved later than other Haps (Supplemental Figure 12B and 12C). However, Hap2 became the dominant haplotype (90%) in KSR after domestication (Supplemental Figure 12D and 12E). Considering the significance of Hap2 in KSR, we speculated that it might be the domesticated haplotype of *Hd1*. It appears that *Hd1* underwent breeding pressure and became a target gene at the stage of KSR breeding selection; that is, compared with APs, KSR matured earlier (Supplemental Figure 12F). RNA-seq and quantitative real-time PCR results showed that the expression of *Hd1* in early-maturing KSR varieties was significantly higher than that in APs (Supplemental Figure 12G and 12H). The above results provide strong genomic evidence for our ethnobiological investigation: compared with APs, KSR had an earlier heading date, enabling it to adapt to the insufficient light and cold cultivation conditions in the mountainous environment of QDN (Huang and Lai, 1983). In addition, *GS3* (Fan et al., 2006), a major quantitative trait locus for grain length on chromosome 3, was also selected (Figure 4A, 4B, and 4C); it functions as a negative regulator of grain length (Mao et al., 2010). The haplotypes of all KSRs were of the domesticated type (Figure 4D, 4E, and 4F), and Hap3 was superior (Figure 4G). Although Hap3 evolved later than other haplotypes (Figure 4D and 4E), the frequency of this superior haplotype extended to 85% of the KSR population (Figure 4F), implying a breeding selection strategy of the Dong people, because large and round grain varieties are required by their traditional dietary culture (Wang et al., 2018a) (Figure 4H). The RNA-seq data showed that *GS3* had higher expression in KSR than in APs (Figure 4I), and quantitative real-time PCR showed that the expression of *GS3* was also more abundant in KSR (Figure 4J). Compared with APs, KSR had a rounder, fuller grain, with increased thousand-grain weight and reduced length-to-width ratio, consistent with a previous study (Li et al., 2019) (Figure 4H and 4K). This was a requirement for the traditional dietary culture of the Dong people. Also, many other genes associated with plant architecture traits were selected. These included *D-h* (Piao et al., 2014), *OsLIC1* (Zhu et al., 2021), and *FC1* (Li et al., 2008). We also identified rice blast and brown planthopper resistance genes such as *Pish* (Imbe and Matsumoto, 1985) and others.

Genome-wide SNP analysis also allowed us to examine the subsequent adaptation of KSR to local environments. One key trait associated with geographical adaptation in KSR is cold tolerance. Based on an ethnobiological survey, most KSR terraces are hidden in the forest area around cottages. This is due to the complex terrain in QDN, which includes a great deal of mountainous area and little or no flat land (Liu et al., 2022a). Many KSR varieties therefore have strong cold tolerance. We interviewed 229 Dong farmers in 10 QDN villages about the mitigation of cold stress damage to plants. The major strategy was to farm cold-tolerant varieties. As expected, 18 genes involved in the regulation of cold tolerance were determined by Pi, F_{ST}, or XP-CLR values to have been selected (Supplemental Figure 13). Analyses of these genes showed unique haplotypes of *LTG1* and *MYBS3* in KSR. *LTG1* was cloned from cold-tolerant *japonica* by Lu et al. (2014). This gene encodes casein kinase I, which is involved in indole-3-acetic acid synthesis and positively regulates low-temperature tolerance of rice during the vegetative growth stage. Five haplotypes were identified;

(E) Evolution network of the main haplotypes of *GS3*.
(F) Geographic distribution and frequency changes of *GS3* haplotypes.
(G) Boxplot of grain LW of different *GS3* haplotypes. Rice landraces were planted in Sanya in 2020 (left) and in QDN in 2021 (right); the asterisks indicate significant differences (*$P < 0.05$; **$P < 0.01$).
(H) Seed-size display of KSR, GZ, and AP populations.
(I) Transcriptomic patterns of *GS3* in KSR and AP, based on the number of fragments per kilobase of the exon model per million mapped reads (FPKM).
(J) Comparison of *GS3* expression levels between KSR and AP by quantitative real-time PCR. Error bars indicate ± SD; $n = 3$ independent biological replicates.
(K) Boxplot of thousand-grain weight of five typical test accessions. Rice landraces were planted in Sanya in 2020 (left) and in QDN in 2021 (right); the asterisks indicate significant differences (**$P < 0.01$).

notably, the superior haplotype Hap3 dominated in KSR (56%) (Supplemental Figure 14A, 14B, and 14C), and it displayed a higher value for *D* (the comprehensive evaluation parameter for cold resistance) and a lower leaf withering degree (LWD) value than other haplotypes (Supplemental Figure 14D). In tracing the evolutionary origin of Hap3 in *LTG1*, we found that KSR was derived from SC in Hap3 (Supplemental Figure 14A). This result suggested that Hap3 was present in SC and was further maintained and extended during KSR domestication through natural and artificial selection. *LTG1* was expressed at higher levels in KSR than in AP varieties when plants were incubated at 10°C for 4 h, 8 h, or 12 h (Supplemental Figure 14E). The results of haplotype, phenotype, and gene expression analyses of *MYBS3* (which plays a critical role in cold adaptation in rice; Su et al., 2010) were similar to those of *LTG1* (Supplemental Figure 15).

KSR was also found to be glutinous, and the pericarp color was variable (white, red, or brown). Haplotype analysis was performed for *Waxy* (Wang et al., 1990) and genes associated with pericarp color (e.g., *Ra* and *Rd*) (Furukawa et al., 2006; Gross et al., 2010) in the KSR, EC, SC, CC, SW, and GZ populations. We found that KSR, like other rice landraces, had no specific haplotypes, suggesting that these traits were selected by humans in the AP. However, haplotype analysis of glutinous rice, non-glutinous rice, and rice with white or colored pericarps showed differences in the haplotype frequencies of these regulatory genes (Supplemental Figure 16).

Genome-wide association study reveals the genetic basis of agronomic and cold-tolerance traits of KSR

Using a genome-wide association study (GWAS), we identified 138 loci that were significantly associated with eight agronomic traits and one cold-tolerance trait (Figure 5 and Supplemental Table 13). Among them, 80 significant loci overlapping with known functional genes were successfully identified in these GWAS peaks, such as *FC1* (Li et al., 2008), *qHd1* (Chen et al., 2014), *GLW10* (Yuan et al., 2022), *EP3* (Piao et al., 2009), and *GS3* (Fan et al., 2006) (Figure 4C). In addition, the intersections of regions subject to selection with GWAS peaks revealed loci that may be related to desirable traits promoted during domestication of KSR. For example, a series of genes with known functions was successfully identified in both GWAS peaks and selection sweeps, including *OsKS1* (Marcia et al., 2005), *OsKS2* (Ji et al., 2014), and *FC1* for panicle length; *Se13* (Xu et al., 2013) for grain number per panicle; *OsCYP51G3* (Xia et al., 2015) for seed-setting rate; *GS3* for grain length-to-width ratio; *EP3* for thousand-grain weight; and *DTH3* (Bian et al., 2011) for days to heading; all of these genes were selected in GZ vs. KSR. Furthermore, *OsMADS51* (Kim et al., 2007) (for panicle length) and *EP3* (for thousand-grain weight) were selected in the AP vs. KSR group.

Some new and highly promising associations were also identified. These included *Os02g0202900* (OsFBK3), an F-box domain-containing protein that regulates seed size and grain number in rice (Chen et al., 2013), and *Os09g0471100*, a peroxidase precursor associated with auxin metabolism and cell-wall thickening (Passardi et al., 2005); both were discovered by GWAS for thousand-grain weight. These genotypes contributed significantly to phenotypic variation (Supplemental Figure 17A, 17B, 17E, and 17F). Both of these genes were expressed at higher levels in KSR compared with AP varieties (Supplemental Figure 17C and 17G). Furthermore, we also used quantitative real-time PCR to investigate the expression of *Os02g0202900* and *Os09g0471100* and found that they were more abundant in KSR than in AP accessions (Supplemental Figure 17D and 17H). In addition, *Os01g0923300* (CBS) was identified in GWAS for grain number per panicle. CBS is a cystathionine β-synthase domain-containing protein that is required for anther and panicle development in rice (Zafar et al., 2020). The frequency of Hap1 of *Os01g0923300* was nearly 90% in KSR, and the phenotype of Hap1 was significantly superior to that of Hap2 (Supplemental Figure 17I and 17J). Hap1 of *Os01g0923300* was therefore determined to be specific to KSR and is worthy of further study. *Os01g0923300* was expressed at higher levels in KSR than in AP varieties (Supplemental Figure 17K), as verified via quantitative real-time PCR (Supplemental Figure 17L). Combined with the plant phenotypes, high expression of these genes in the spikelet primordium indicated that they function in regulating rice grain and spikelet development.

DISCUSSION

In this study, we present a large-scale survey of rice landrace genomes from southern China that cover all major areas of the rice-farming civilization in the region. Using genomic data, we revealed the evolutionary history of these accessions and hypothesize about the historical migration routes of the Dong people. This approach is scientifically sound because KSR is cultivated and utilized only by the Dong people and because they carried KSR during their migrations (Yang, 2021; Liu et al., 2022a). In addition, the KSR samples used in this study were collected in 1980, before the introduction of other rice landraces or hybrid rice lines from surrounding provinces (Ruan et al., 2007). Thus, there was almost no recent gene introgression in these KSR samples. This effective and ingenious approach avoids the drawbacks of insufficient quantitative data faced by archaeological and ethnological studies of the Dong people (Chen and Wei, 2018). It also circumvents the problems of few sample materials, poor sample integrity, difficulties in extraction of ancient DNA, and vulnerability to external contamination encountered in the study of ancient human genomes (Lipson et al., 2018).

Our population genomics study showed that KSR originated from rice landraces in Guangxi, Guangdong, Fujian, and Jiangxi provinces. KSR was brought to QDN when the ancestors of the Dong people migrated, supporting Route III as proposed by historians, archaeologists, and ethnologists. Combined with current historical research (Fan, 1989; Xu and Li, 2014), our results showed that the ancestors of the Dong people migrated multiple times (Figure 6). The distribution of the Dong people is a result of the genetic origin, migration, and admixture history of this group, which is consistent with research conducted by Ma et al. on the origin of the Hui people (Ma et al., 2021). Consistent with the results discussed above, we found mixed patterns of population structure and high frequencies of genetic exchange between KSR and SC or EC populations. Such frequent gene flow in geographically non-adjacent regions implies human

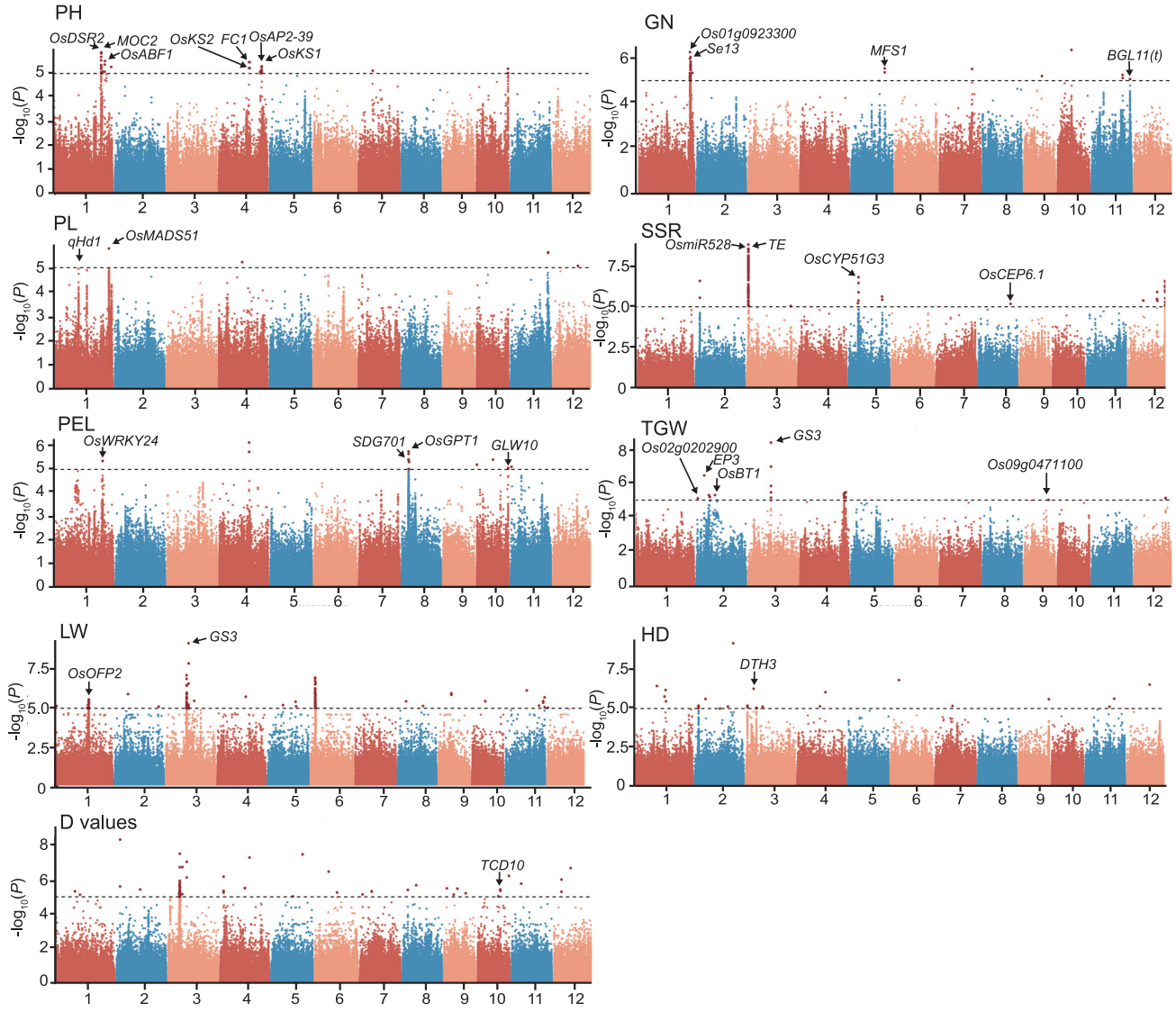

Figure 5. Manhattan plots of GWAS for eight agronomic traits and one cold-tolerance trait ($-\log_{10}P = 5$).
These traits comprise plant height (PH), panicle length (PL), peduncle length (PEL), grain number per panicle (GN), seed-setting rate (SSR), thousand-grain weight (TGW), grain length-to-width ratio (LW), days to heading (HD), and cold-tolerance *D* value. Dashed lines indicate the threshold for GWAS ($-\log_{10}P = 5$). Candidate genes for GN and TWG closest to the GWAS signals are displayed.

intervention (Zeng et al., 2018). SC and EC are donors to KSR, as illustrated by the similar haplotype patterns observed from some representative introgressions in these three populations (Wu et al., 2022). Thus, gene flow due to human activity may be the primary factor contributing to the observed patterns of KSR domestication, consistent with results in other crops such as sorghum (Wu et al., 2022), wheat (Zhou et al., 2020), and castor bean (Xu et al., 2021). We also found that there has been genetic differentiation between KSR in QDN and the ancestral rice populations of SC and EC. This divergence was estimated to have occurred ~350 YBP. Early evidence shows that the Dong people settled as a single ethnic community in QDN 1000 years ago during the Song Dynasty (Yang, 2014). We therefore infer that the Dong people have been domesticating KSR for ~650 years. As an ecotypic rice landrace with characteristics unique to QDN, KSR has adapted to the local natural environment and the traditional culture of the Dong people and has been passed down through the generations.

Characterization of genes associated with domestication could shed light on the process of crop domestication at the genetic level (Wu et al., 2022), in which artificial selection plays a key role (Olsen and Wendel, 2013). In KSR, we found evidence for selection of domestication genes and identified the models of haplotype changes in three key genes (*Hd1*, *GS3*, and *DPS1*) from APs to KSR. These results provide strong genetic evidence for a dynamic domestication process of KSR by the Dong people. Previous studies have shown that *Hd1* has undergone human selection to diversify rice flowering time during domestication (Takahashi and Shimamoto, 2011). We confirmed this finding and further dissected the haplotype evolutionary network of *Hd1* in different populations. This

Figure 6. Proposed migration routes of the Dong people reconstructed based on genetic evidence from KSR and historical records.
The pink arrows and inset photographs show the domestication process from the ancestral population to KSR. The red arrows indicate proposed migration routes of the Dong ancestors reconstructed from genetic evidence for KSR and historical records.

haplotype analysis revealed changes in heading date during KSR domestication by the Dong people. They selected for earlier heading in KSR to minimize reductions in grain yield caused by low temperature, confirming our ethnological findings at the genetic level. In addition, significant evolution of agronomic traits was detected in KSR. The panicle length, grain number per panicle, seed-setting rate, and thousand-grain weight were higher in KSR than in the AP. This is consistent with the dietary and economic needs of the Dong people and suggests their positive input during domestication. In particular, the role of *GS3* in regulating rice grain size deserves an in-depth investigation into the associated complex regulatory network (Sun et al., 2018). KSR has a full, rounded grain phenotype, which is clearly different from that of the AP and other rice landraces in Guizhou. This may be because the expression level of *GS3* is precisely regulated in each accession. Our GWAS results also established a link between grain size and *GS3*. In addition, we found that glutinous traits and a colored pericarp were not further selected by the Dong people, and the lines with red or brown pericarps were present in their ancestral genetic pool.

We also detected genetic change that was related to environmental changes associated with migration of the Dong people (Xu et al., 2021), such as identification of the cold-tolerance regulatory genes *LTG1* and *MYBS3*. Haplotype analysis further revealed that KSR had a higher frequency of the superior haplotypes of these genes, suggesting that they were targets during KSR domestication. The phylogenetic tree and network of haplotypes also revealed genetic evolution of KSR that was consistent with the migration route of the Dong people and showed that superior haplotypes expanded to become dominant in the KSR population after domestication. Owing to environmental changes, the Dong people have selected varieties with strong cold tolerance as an agronomic strategy (Liu et al., 2022b). Through the generations, KSR cold tolerance has gradually been enhanced, and genomic footprints have been left that reveal profound changes in the genome due to selection. All of these findings are consistent with and validate our ethnobiological survey results from a genomic perspective.

Combining a genome scan with GWAS allowed us to investigate the genetic basis of agronomic traits, especially those related to heading date, seed size, and seed-setting rate, which were clearly targets of selection by the Dong people. Furthermore, we identified three possible candidate genes; both RNA-seq and quantitative real-time PCR verified that expression levels of the newly identified genes were significantly higher in the KSR population than in the AP. This indicates a need for further study of genes regulating yield traits.

Although the migration of the Dong people occurred long ago, the significance and particularity of KSR in Dong agricultural production mean that we can use genetic footprints in KSR to identify the migration route of the Dong people. In the future, the complementation and deep integration of natural sciences and social sciences is an important way to explore the evolution of human civilizations. Our results suggest a scenario of origin, migration, and admixture history in the Dong population through a multidisciplinary approach. This study not only generates new insights into the migration of the Dong people by genetic evolution and domestication of KSR but also presents new tools for the study of human migration.

METHODS

Ethnobiological survey

We conducted continuous ethnobiological studies in Dong villages in Guizhou Province in September of 2015, July of 2019, August of 2020, and October of 2021. Two hundred twenty-nine respondents (Dong farmers) were interviewed through the snowball technique, which included key informant interviews, semi-structured interviews, and participatory observations. Survey sites included farmers' homes, farmlands, fish ponds, village roads, and workshops. Key informant interviews were conducted primarily with local experts, village cadres, clan elders (e.g., village elders or headmen), and inheritors of intangible cultural heritage. The semi-structured interviews involved open-ended questions and conversations with informants in the locations listed above. All interview procedures were in accordance with the International Society of Ethnobiology code of ethics, including procuring informed consent prior to interviews (http://ethnobiology.net/code-of-ethics/). We also conducted participatory observations; specifically, we devoted ourselves to observation of the life of local ethnic minorities, focusing on the role of KSR in the daily life and traditional culture of the Dong people, the layout of KSR farmland ecosystems, traditional KSR management, and the daily production and collective activities of the Dong people. This allowed us to better understand the Dong people.

Sampling, DNA isolation, and genome sequencing

In this study, a total of 372 core rice landraces were selected from 1481 accessions. We divided the 372 core accessions into six groups. KSR comprised 104 KSR accessions from QDN. GZ included 104 rice landraces from Guizhou (except QDN). There were 23 landraces from central China (CC), 74 from East China (EC), 36 from South China (SC), and 31 from Southwest China (SW). Details of accessions are shown in Supplemental Table 1 and Supplemental Figure 3. The selected accessions represented a wide range, including all possible APs of KSR.

Genomic DNA was extracted using standard methods, and 1.5 μg of DNA per sample was used as input for sequencing library generation using a TruSeq Nano DNA HT sample preparation kit (Illumina, USA) following the manufacturer's recommendations. Index codes were added to attribute sequences to each sample. In brief, the libraries were prepared as follows: the genomic DNA sample was fragmented by sonication to a size of ∼350 bp, then DNA fragments were end-polished, A-tailed, and ligated with the full-length adapters for Illumina sequencing with further PCR amplification. PCR products were then purified using the AMPure XP bead system. The libraries were analyzed for size distribution using an Agilent 2100 Bioanalyzer and quantified by real-time PCR. Subsequently, we used the Illumina NovaSeq platform to generate ∼2.06 Tb of raw sequences with a 150-bp read length.

Sequence quality checking and filtering

To eliminate reads with artificial bias (i.e., low-quality paired reads, which primarily result from base-calling duplicates and adaptor contamination), we removed the following types of reads: (i) reads with ≥10% unidentified nucleotides; (ii) reads with >10 nt aligned to the adaptor, with ≤10% mismatches allowed; (iii) reads in which >50% bases had Phred quality scores <5; and (iv) putative PCR duplicates generated via PCR amplification in the library construction process (i.e., read 1 and read 2 of two paired-end reads that were completely identical). Consequently, we retained 1.99 Tb (∼5.97 Gb per sample) of high-quality genomic data (Supplemental Table 3 and 4).

Sequence alignment, variation calling, and annotation

The remaining high-quality paired-end reads were mapped to the *O. sativa japonica* reference genome (IRGSP-1.0) using Burrows-Wheeler Aligner software (Li and Durbin, 2009) with the command "mem -t 4 -k 32 -M". To reduce mismatches generated by PCR amplification before sequencing, duplicated reads were removed using SAMtools (v.0.1.1) (Li et al., 2009). After alignment, the genomic variants (in GVCF format for each line) were identified using the Sentieon DNASeq software package (Weber et al., 2016). Then all GVCF files of all lines were merged. To obtain high-quality variants, variants with QD > 2.0, FS < 60.0, MQ > 20.0, MQRankSum > −12.5, and ReadPosRankSum > −8.0 were retained using GATK software (McKenna et al., 2010). The following potential low-quality variants were removed: (1) two alleles only, (2) missing rate <0.2, (3) minor allele individuals ≥5.0, and (4) minor allele frequency ≤0.05. Consequently, a total of 3 566 872 SNPs were retained. The identified SNPs and indels were further annotated with ANNOVAR software (version 2013-05-20) (Wang et al., 2010) and thus divided into the following groups on the basis of newly updated rice genome annotation information: variations in intergenic regions, within 1 kb upstream (downstream) of transcription start (stop) sites, in coding sequences, and in introns.

Phenotyping

The 397 rice landraces were planted and phenotyped in Sanya of Hainan province and QDN of Guizhou Province in 2020 and 2021, respectively. At least 10 plants in the middle of the plot were selected for phenotyping; measurements included plant height (cm), panicle length (cm), peduncle length (cm), grain number per panicle, seed-setting rate, thousand-grain weight (g), grain length-to-width ratio, and days to heading (days), according to the standard evaluation system for rice (Han and Wei, 2006).

Phylogenetic tree and population structure

We used 3 566 872 SNPs to construct the individual-based neighbor-joining tree based on the p distance using TreeBest software (v.1.9.2) (Vilella et al., 2005) with 1000 bootstrap replications. The population genetic structure was examined using ADMIXTURE (v.1.23) (Alexander et al., 2009). PCA was performed using GCTA software (Yang et al., 2011). First, we obtained the genetic relationship matrix with the parameter "–make-grm". Then, the top three principal components were estimated with the parameter "–pca3".

Population genetic analysis

Nucleotide diversity π for each group (KSR, CC, EC, SC, SW, GZ, WR, Bas, and Aus) and fixation index (F_{ST}) between populations were analyzed by VCFtools (v.0.1.14) (Danecek et al., 2011) within a 20-kb window.

Introgression analyses

To detect introgression events from other populations into KSR, we used TreeMix version 1.13 (Pickrell and Pritchard, 2012) to investigate gene flow between groups with the settings "-tf -se -bootstrap 8 -k 1000 -root -m", where the number (-m) varied from 1 to 8.

To verify the introgression events detected by TreeMix, we applied the ABBA–BABA test (*D* statistic) by detecting differences in allele sharing between two lineages (P1 and P2) with a third lineage (P3) (Durand et al., 2011). Wild rice was used as the outgroup and SW as the P1 group, and a significant gene flow signal was defined as a *D* statistic larger than 3.

Subsequently, the f_d statistic (Martin et al., 2014) was computed to identify introgressed genomic fragments based on the tree form (((P1, P2), P3), O), consistent with the above *D* statistic. The f_d statistic was computed in 20-kb non-overlapping windows with the python script ABBABABAwindows.py (https://github.com/simonhmartin/genomics_general).

We inferred fluctuations in effective population size for KSR, EC, SC, and WR with SMC++ (v.1.15.2) (Terhorst et al., 2017) based on a constant generation time of 1 year and a per-generation mutation rate of 6.5×10^{-9} (Gaut et al., 1996).

Selective signals of adaptation

To identify potential selective signals during KSR evolution, we used a sliding-window approach (20-kb windows sliding in 10-kb steps) to quantify the levels of nucleotide diversity (π) and genetic differentiation (F_{ST}) between KSR and GZ or AP groups using VCFtools (v.0.1.14) (Danecek et al.,

2011). The π ratios between groups were log transformed. We also used the XP-CLR score to scan for domestication-sweep regions (–w1 0.005 200 2000 1 –p0 0.95) (Chen et al., 2010). The windows with the top 5% of F_{ST}, π ratio (π_{CL}/π_{MCC}), or XP-CLR scores were considered to be candidate sweeps during breeding.

KEGG pathway analysis

Enriched rice pathways (http://www.kegg.jp) in candidate genes selected during KSR evolution were identified using KOBAS v.2.0 (Mao et al., 2005). The filtering criteria of $P < 0.05$ and FDR < 0.05 were used for the KEGG enrichment analysis.

Cold-tolerance phenotype evaluation

To examine variation in cold tolerance, we measured several cold-associated fitness traits in 209 typical *japonica* landraces at the seedling stage. Tested traits were LWD, chlorophyll content, nitrogen content, seedling length, root length, and plant dry weight. These parameters were measured as described by Han and Wei (2006) and Han et al. (2022).

Ten seeds per accession were sown in seedling trays in August 2021 in Beijing. Seedlings were incubated at 28°C prior to low-temperature treatment. Starting at the 3.5-leaf stage, plants in the low-temperature group were incubated at 10°C for 7 days to mimic cold stress. Seedlings were then transferred to 28°C for 7 days to recover. The control group was incubated for the entire time at 28°C. LWD was measured at day 7 of treatment and at day 7 of recovery. LWD values ranged from 1 to 9, with 1 indicating the strongest possible cold tolerance (all plants survived) and 9 indicating the weakest possible cold tolerance (all plants died). Chlorophyll content and nitrogen content were measured with three technical replicates using a plant nutrition tester (TYS-4 N; Top YunNong, Zhejiang) at three test points that were 3–5 cm from the leaf tip. Seedling length and root length were measured in three plants per accession using a ruler. After plants were measured, they were cleaned with gauze, wrapped in newspaper, dried at 105°C for 30 min, and then dried at 80°C to a constant weight. After cooling, dry weight was measured with an analytical balance. The average value was calculated for each parameter.

To eliminate the differences in basic characteristics between different varieties, relative values of the measurement indexes were used to evaluate cold tolerance as follows:

$$Relatives\ value = LT/CG. \quad \text{(Equation 1)}$$

Cold resistance was analyzed using subordinate function methods, and x_j was calculated with PCA using SPASS 24.0 software. The subordinate function $\mu(x_j)$ was calculated as follows:

$$\mu(x_j) = (x_j - x_{min}) / (x_{max} - x_{min}) \times 100\%, \quad \text{(Equation 2)}$$

where x_j is the ith comprehensive index and x_{min} and x_{max} are the minimum and maximum, respectively, for the ith comprehensive index.

The weight function w_j was calculated and represents the relative importance of the ith comprehensive index:

$$w_j = r_j \Big/ \sum_{j=1}^{n} r_j, \quad \text{(Equation 3)}$$

where γ_j represents the contribution for the ith comprehensive index.

The comprehensive evaluation parameter for cold resistance D was calculated as follows:

$$D = \sum_{j=1}^{n} [\mu(x_j) w_j]. \quad \text{(Equation 4)}$$

D and LWD were selected as the final evaluation indexes for cold resistance in this study. Higher D values and lower LWD values were associated with stronger cold tolerance.

Genome-wide association study

We performed a GWAS for eight agronomic traits and one cold-tolerance trait: plant height, panicle length, peduncle length, grain number per panicle, seed-setting rate, thousand-grain weight, grain length-to-width ratio, days to heading, and cold-tolerance D value. Two hundred nine *japonica* accessions were used to perform GWAS using the GEMMA (genome-wide efficient mixed-model association) software package (Zhou and Stephens, 2012). For the mixed-linear-model analysis we used the following equation:

$$y = X\alpha + S\beta + K\mu + e,$$

where y represents phenotype; α and β are fixed effects representing marker effects and non-marker effects, respectively; and μ represents unknown random effects. X, S, and K are the incidence matrices for α, β, and μ, respectively, and e is a vector of random residual effects. The top three PCs were used to build up the S matrix for population-structure correction. The matrix of simple matching coefficients was used to build up the K matrix. The analyses were performed in the GEMMA software package. Genetic relationship between individuals was modeled as a random effect using the kinship (K) matrix. Significant P thresholds ($P < 10^{-5}$) were set to control the genome-wide type I error rate. Then we enlarged the candidate region to 100 kb centered on the GWAS signal peak to identify candidate genes.

Expression analysis with RNA-seq

The cold-tolerance test was performed using six selected accessions: three KSR populations and three APs. Seeds of all accessions were sown in equal-size seedling trays for three replicates in March 2022 in Beijing. Ten seeds were used per accession, and seedlings were incubated at 28°C. When the seedlings had grown to the 3.5-leaf stage, they were incubated at 10°C to mimic temperature stress. At 4 h, 8 h, and 12 h after cold stress, leaf tissue from the flag leaf center was collected and frozen in liquid nitrogen. Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen) following the manufacturer's instructions for sequencing analysis. Sequencing was performed on the Illumina X Ten platform. Approximately 6 Gb of 150-bp paired-end reads were generated for each sample. The clean reads from these RNA-seq experiments were mapped onto the *O. sativa japonica* reference genome (IRGSP-1.0) using TopHat2 (v.2.1.1) (Trapnell et al., 2012). The expression level of each transcript was counted and normalized into fragments per kilobase per million mapped reads using Cufflinks (v.2.1.1) (Pachter et al., 2010).

RNA extraction and quantitative real-time PCR

Total RNA was extracted from spikelet primordia of KSR and ancestral samples using a TIANGEN kit (Tiangen Biotech) for quantitative real-time PCR analysis. Approximately 500 ng of total RNA was used to synthesize first-strand complementary DNA with RT-gDNA digestion SuperMix for the qPCR kit (Yeason Biotechnology). The quantitative real-time PCR was performed using ChamQ SYBR Color qPCR Master Mix Q411 (Vazyme Biotech). The gene encoding ubiquitin-60S ribosomal protein (*Os03g0234200*) served as an internal control. Primers for the candidate genes were designed using Primer Premier 5. Primers used in this analysis are listed in Supplemental Table 14.

Data availability

The raw sequencing data have been deposited in the China National Center for Bioinformation (https://ngdc.cncb.ac.cn/gsa/) under accession code PRJCA010430. Other data used are listed in Supplemental Table 1.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Molecular Plant Online*.

FUNDING

This work was supported by the National Key Research and Development Program of China (2021YFD1200500), the National Natural Science Foundation of China (31901487), the CAAS Science and Technology Innovation Program, the Protective Program of Crop Germplasm of China (19200385-1), the Third National Survey and Collection Action on Crop Germplasm Resource (19210859, 19210860), and the National Crop Germplasm Resources Center (NCGRC-2021-02).

AUTHOR CONTRIBUTIONS

C.H.L., as the principal researcher, conducted the experiments and data analysis and drafted the manuscript. Y.J.W. and L.Z.H., as supervisors, initiated the study; they provided input into its planning, conducted oversight of the study, and offered ethical and cultural advice based on their rich experience and knowledge. T.Y.W. and C.Z.J. assisted in analyzing the data. C.H.L., H.C.C., X.B.L., A.X.J., and R.C.R. carried out the field planting and phenotypic survey. C.H.L., Y.J.W., D.Y.X., and L.Z.H. conducted ethnobiological studies. X.D.M., D.C., and B.H. carefully revised the manuscript.

ACKNOWLEDGMENTS

We are grateful to the Dong farmers for help with the ethnobiological survey, and the portraits of Dong people used in this article are all used with permission. We thank photographer Wei Wang, who helped us take some photos in the Dong village. We thank the National Medium-Term Gene Bank, Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, for providing the rice seeds. We are very grateful to Professor Zhengyi Xu of Northeast Normal University of China for help with the haplotype analyses. No conflict of interest is declared. The authors declare that they have no conflict of interest.

Received: July 17, 2022
Revised: November 22, 2022
Accepted: December 19, 2022
Published: December 27, 2022

REFERENCES

Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Res. **19**:1655–1664.

Bedigian, D. (2003). Specialty rices of the world. Breeding, production and marketing. Econ. Bot. **57**:160.

Bian, X.F., Liu, X., Zhao, Z.G., Jiang, L., Gao, H., Zhang, Y.H., Zheng, M., Chen, L.M., Liu, S.J., Zhai, H.Q., et al. (2011). Heading date gene, dth3 controlled late flowering in O. Glaberrima Steud. by down-regulating Ehd1. Plant Cell Rep. **30**:2243–2254.

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. Genome Res. **20**:393–402.

Chen, H.L., and Wei, D.F. (2018). Review and reflection on Chinese ethnoarchaeology research in the last 30 years. J. Guangxi University for Nationalities **40**:25–34.

Chen, J.Y., Guo, L., Ma, H., Chen, Y.Y., Zhang, H.W., Ying, J.Z., and Zhuang, J.Y. (2014). Fine mapping of qHd1, a minor heading date QTL with pleiotropism for yield traits in rice (Oryza sativa L.) Theor. Theor. Appl. Genet. **127**:2515–2524.

Chen, Y., Xu, Y., Luo, W., Li, W., Chen, N., Zhang, D., and Chong, K. (2013). The F-Box Protein OsFBK12 Targets OsSAMS1 for degradation and affects pleiotropic phenotypes, including leaf senescence, in rice. Plant Physiol. **163**:1673–1685.

Cui, D., Tang, C., Lu, H., Li, J., Ma, X., A, X.X., A, X., Han, B., Yang, Y., Dong, C., et al. (2021). Genetic differentiation and restricted gene flow in rice landraces from Yunnan, China: effects of isolation-by-distance and isolation-by-environment. Rice **14**:54.

Cui, D., Zhou, H., Ma, X., Lin, Z., Sun, L., Han, B., Li, M., Sun, J., Liu, J., Jin, G., et al. (2022). Genomic insights on the contribution of introgressions from Xian/Indica to the genetic improvement of Geng/Japonica rice cultivars. Plant Commun. **3**:100325.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., Depristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. Bioinformatics **27**:2156–2158.

Delteil, A., Gobbato, E., Cayrol, B., Estevan, J., Michel-Romiti, C., Dievart, A., Kroj, T., and Morel, J.B. (2016). Several wall-associated kinases participate positively and negatively in basal defense against rice blast fungus. BMC Plant Biol. **16**:17.

Du, X., Huang, G., He, S., Yang, Z., Sun, G., Ma, X., Li, N., Zhang, X., Sun, J., Liu, M., et al. (2018). Resequencing of 243 diploid cotton accessions based on an updated a genome identifies the genetic basis of key agronomic traits. Nat. Genet. **50**:796–802.

Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. Mol. Biol. Evol. **28**:2239–2252.

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., Li, X., and Zhang, Q. (2006). GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor. Appl. Genet. **112**:1164–1171.

Fan, H.G. (1989). Place and time of migration of Dongs' ancestors. Guangxi ethnic Studies **4**:62–67.

Fuller, D.Q., Sato, Y.I., Castillo, C., Qin, L., Weisskopf, A.R., Kingwell-Banham, E.J., Song, J., Ahn, S.M., and van Etten, J. (2010). Consilience of genetics and archaeobotany in the entangled history of rice. Archaeol. Anthropol. Sci. **2**:115–131.

Furukawa, T., Maekawa, M., Oki, T., Suda, I., Iida, S., Shimada, H., Takamure, I., and Kadowaki, K.i. (2006). The *Rc* and *Rd* genes are involved in proanthocyanidin synthesis in rice pericarp. Plant J. **49**:91–102.

Gaut, B.S., Morton, B.R., McCaig, B.C., and Clegg, M.T. (1996). Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene Adh parallel rate differences at the plastid gene rbcL. Proc. Natl. Acad. Sci. USA. **93**:10274–10279.

Gaut, B.S., Seymour, D.K., Liu, Q., and Zhou, Y. (2018). Demography and its effects on genomic variation in crop domestication. Nat. Plants **4**:512–520.

Ge, J.X. (2005). History of Population in China (Shanghai: Fudan University Press).

Gross, B.L., Steffen, F.T., and Olsen, K.M. (2010). The molecular basis of white pericarps in African domesticated rice: novel mutations at the Rc gene. J. Evol. Biol. **23**:2747–2753.

Han, B., Cui, D., Ma, X., Cao, G., Zhang, H., Koh, H.J., and Han, L. (2022). Evidence for evolution and selection of drought-resistant genes based on high-throughput resequencing in weedy rice. J. Exp. Bot. **73**:1949–1962.

Han, L.Z., and Wei, X.H. (2006). Descriptors and Data Standard for Rice (*Oryza sativa* L.) (Beijing: China Agricul. Press).

Haudry, A., Cenci, A., Ravel, C., Bataillon, T., Brunel, D., Poncet, C., Hochu, I., Poirier, S., Santoni, S., Glémin, S., et al. (2007). Grinding up wheat: a massive loss of nucleotide diversity since domestication. Mol. Biol. Evol. **24**:1506–1517.

He, Y.T. (2011). A study on disasters in Guizhou Province in modern times. J. Changjiang Tribune **6**:61–68.

He, Z., Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Zhang, L., He, W., Lu, B., Lin, H., et al. (2008). Control of rice grain-filling and yield by a gene with a potential signature of domestication. Nat. Genet. **40**:1370–1374.

Huang, G.Q. (2016). Planting, food making and sacrifice giving with glutinous rice among ethnic groups of the Dong linguistic family in China. Guangxi Ethnic Studies **1**:80–87.

Huang, W.X., and Lai, J.H. (1983). Characteristics of light and temperature response in HE resource. Guizhou Agricul. Sci. **2**:15–18.

Imbe, T., and Matsumoto, S. (1985). Inheritance of resistance of rice varieties to the blast fungus strains virulent to the variety "reiho". Ikushugaku Zasshi **35**:332–339.

Ji, S.H., Gururani, M.A., Lee, J.W., Ahn, B.O., and Chun, S.C. (2014). Isolation and characterisation of a dwarf rice mutant exhibiting defective gibberellins biosynthesis. Plant Biol. **16**:428–439.

Kim, S.L., Lee, S., Kim, H.J., Nam, H.G., and An, G. (2007). OsMADS51 is a short-day flowering promoter that functions upstream of Ehd1, OsMADS14, and Hd3a. Plant Physiol. **145**:1484–1494.

Kutanan, W., Liu, D., Kampuansai, J., Srikummool, M., Srithawong, S., Shoocongdej, R., Sangkhano, S., Ruangchai, S., Pittayaporn, P., Arias, L., et al. (2021). Reconstructing the human genetic history of mainland Southeast Asia: insights from genome-wide data from Thailand and Laos. Mol. Biol. Evol. **38**:3459–3477.

Larson, G., Piperno, D.R., Allaby, R.G., Purugganan, M.D., Andersson, L., Arroyo-Kalin, M., Barton, L., Climer Vigueira, C., Denham, T., Dobney, K., et al. (2014). Current perspectives and the future of domestication studies. Proc. Natl. Acad. Sci. USA. **111**:6139–6146.

Lei, Q., Zhou, J.J., Luo, J., Zhang, W.H., Sun, J., and Long, C.L. (2017). Changes in the numbers of Kam fragrant glutinous rice varieties in the Dong regions of Guizhou Province. Biodivers. Sci. **25**:990–998.

Lei, Q.Y., Zhou, J.J., Xiong, Y., Zhang, W.H., Luo, J., and Long, C.L. (2021). Genetic diversity evaluation and conservation of kam fragrant glutinous rice (*Oryza sativa* L.) germplasm in southeast Guizhou, China. Plants **10**:1898.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**:1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079.

Li, J., Zhang, X.F., Huang, G., and Feng, Y.H. (2019). Research progress on local rice varietied of kam sweet rice in southeast of Guizhou. China Rice **25**:53–58.

Li, X., Yang, Y., Yao, J., Chen, G., Li, X., Zhang, Q., and Wu, C. (2008). FLEXIBLE CULM 1 encoding a cinnamyl-alcohol dehydrogenase controls culm mechanical strength in rice. Plant Mol. Biol. **69**:685–697.

Lipson, M., Cheronet, O., Mallick, S., Rohland, N., Oxenham, M., Pietrusewsky, M., Pryce, T.O., Willis, A., Matsumura, H., Buckley, H., et al. (2018). Ancient genomes document multiple waves of migration in southeast Asian prehistory. Science **361**:92–95.

Liu, C., Wang, Y., Ma, X., Cui, D., Han, B., Xue, D., and Han, L. (2022a). Traditional agricultural management of kam sweet rice (*Oryza Sativa* L.) in Southeast Guizhou Province, China. J. Ethnobiol. Ethnomed. **18**:30.

Liu, C., Cui, D., Jiao, A., Ma, X., Li, X., Han, B., Chen, H., Ruan, R., Wang, Y., and Han, L. (2022b). Kam sweet rice (*Oryza sativa* L.) is a special ecotypic rice in southeast Guizhou, China as revealed by genetic diversity analysis. Front. Plant Sci. **13**:830556.

Lu, G., Wu, F.Q., Wu, W., Wang, H.J., Zheng, X.M., Zhang, Y., Chen, X., Zhou, K., Jin, M., Cheng, Z., et al. (2014). Rice *LTG1* is involved in adaptive growth and fitness under low ambient temperature. Plant J. **78**:468–480.

Ma, H.L. (1979). HE Resource and Cultivation of HE Resource//.HE Resource and Agricultural Production (Guiyang: Compiled by the office of crop germplasm resources of Guizhou Province).

Ma, X., Yang, W., Gao, Y., Pan, Y., Lu, Y., Chen, H., Lu, D., and Xu, S. (2021). Genetic origins and sex-biased admixture of the Huis. Mol. Biol. Evol. **38**:3804–3819.

Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., Li, X., and Zhang, Q. (2010). Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. Proc. Natl. Acad. Sci. USA. **107**:19579–19584.

Mao, X., Cai, T., Olyarchuk, J.G., and Wei, L. (2005). Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics **21**:3787–3793.

Margis-Pinheiro, M., Zhou, X.R., Zhu, Q.H., Dennis, E.S., and Upadhyaya, N.M. (2005). Upadhyaya. Isolation and characterization of a Ds-tagged rice (*Oryza sativa* L.) GA-responsive dwarf mutant defective in an early step of the gibberellin biosynthesis pathway. Plant Cell Rep. **23**:819–833.

Martin, S.H., Davey, J.W., and Jiggins, C.D. (2014). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Mol. Biol. Evol. **32**:244–257.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**:1297–1303.

Ning, C., Li, T., Wang, K., Zhang, F., Li, T., Wu, X., Gao, S., Zhang, Q., Zhang, H., Hudson, M.J., et al. (2020). Ancient genomes from Northern China suggest links between subsistence changes and human migration. Nat. Commun. **11**:2700.

Oliveira, S., Nägele, K., Carlhoff, S., Pugach, I., Koesbardiati, T., Hübner, A., Meyer, M., Oktaviana, A.A., Takenaka, M., Katagiri, C., et al. (2022). Ancient genomes from the last three millennia support multiple human dispersals into Wallacea. Nat. Ecol. Evol. **6**:1024–1034.

Olsen, K.M., and Wendel, J.F. (2013). A bountiful harvest: genomic insights into crop domestication phenotypes. Annu. Rev. Plant Biol. **64**:47–70.

Pachter, L., Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantifcation by RNA-seq reveals unannotated transcripts and isoform switching during cell diferentiation. Nat. Biotechnol. **28**:511–515.

Pan, Y.R., and Long, X.Y. (2013). Nomenclature and classification of Kam Sweet Rice varieties: a Case study of agricultural cognition of China's mountain peoples. Mountain Peoples and Culture **5**:91–98.

Passardi, F., Cosio, C., Penel, C., and Dunand, C. (2005). Peroxidases have more functions than a Swiss army knife. Plant Cell Rep. **24**:255–265.

Peng, H., Zhang, Q., Li, Y., Lei, C., Zhai, Y., Sun, X., Sun, D., Sun, Y., and Lu, T. (2009). A putative leucine-rich repeat receptor kinase, OsBRR1, is involved in rice blast resistance. Planta **230**:377–385.

Piao, R., Chu, S.H., Jiang, W., Yu, Y., Jin, Y., Woo, M.O., Lee, J., Kim, S., and Koh, H.J. (2014). Isolation and characterization of a dominant dwarf gene, D-h, in rice. PLoS One **9**:e86210.

Piao, R., Jiang, W., Ham, T.H., Choi, M.S., Qiao, Y., Chu, S.H., Park, J.H., Woo, M.O., Jin, Z., An, G., et al. (2009). Map-based cloning of the ERECT PANICLE 3 gene in rice. Theor. Appl. Genet. **119**:1497–1506.

Pickrell, J.K., and Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. **8**:e1002967.

Ruan, R.C., Chen, H.C., You, J.M., Zhu, Y.Q., Zhu, M., and Chen, N.G. (2007). Current status and prospects of conservation for rice genetic resources in Guizhou. Seeds **26**:53–56.

Sato-Izawa, K., Nakaba, S., Tamura, K., Yamagishi, Y., Nakano, Y., Nishikubo, N., Kawai, S., Kajita, S., Ashikari, M., Funada, R., et al. (2012). DWARF50 (D50), a rice (*Oryza sativa* L.) gene encoding inositol polyphosphate 5-phosphatase, is required for proper development of intercalary meristem. Plant Cell Environ. **35**:2031–2044.

Shen, S., Peng, M., Fang, H., Wang, Z., Zhou, S., Jing, X., Zhang, M., Yang, C., Guo, H., Li, Y., et al. (2021). An oryza-specific hydroxycinnamoyl tyramine gene cluster contributes to enhanced disease resistance. Sci. Bull. **66**:2369–2380.

Stone, R. (2008). Intellectual property: Chinese province crafts pioneering law to thwart biopiracy. Science **320**:732–733.

Su, C.F., Wang, Y.C., Hsieh, T.H., Lu, C.A., Tseng, T.H., and Yu, S.M. (2010). A novel MYBS3-dependent pathway confers cold tolerance in rice. Plant Physiol. **153**:145–158.

Sun, S., Wang, L., Mao, H., Shao, L., Li, X., Xiao, J., Ouyang, Y., and Zhang, Q. (2018). A G-protein pathway determines grain size in rice. Nat. Commun. **9**:851.

Takahashi, Y., and Shimamoto, K. (2011). Heading date 1 (Hd1), an ortholog of Arabidopsis CONSTANS, is a possible target of human selection during domestication to diversify flowering times of cultivated rice. Genes Genet. Syst. **86**:175–182.

Terhorst, J., Kamm, J.A., and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. Nat. Genet. **49**:303–309.

Trapnell, C., Roberts, A., Goff, L., PERTEA, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Diferential gene and transcript expression analysis of RNAseq experiments with TopHat and Cufinks. Nat. Protoc. **7**:562–578.

Vilella, A.J., Blanco-Garcia, A., Hutter, S., and Rozas, J. (2005). VariScan: analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. Bioinformatics **21**:2791–2793.

Wang, J., Chen, Z., Zhang, Q., Meng, S., and Wei, C. (2020). The NAC transcription factors OsNAC20 and OsNAC26 regulate starch and storage protein synthesis. Plant Physiol. **184**:1775–1791.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. **38**:e164.

Wang, Q., Su, Q.,M., Nian, J.Q., Zhang, J., Guo, M., Dong, G.J., Hu, J., Wang, R.S., Wei, C.S., Li, G.W., et al. (2021). The *Ghd7* transcription factor represses ARE1 expression to enhance nitrogen utilization and grain yield in rice. Mol. Plant **14**:1012–1023.

Wang, W.S., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., et al. (2018b). Genomic variation in 3010 diverse accessions of Asian cultivated rice. Nature **557**:43–49.

Wang, Y., Jiao, A., Chen, H., Ma, X., Cui, D., Han, B., Ruan, R., Xue, D., and Han, L. (2018a). Status and factors influencing on-farm conservation of Kam Sweet Rice (*Oryza sativa* L.) genetic resources in southeast Guizhou Province, China. J. Ethnobiol. Ethnomed. **14**:76.

Wang, Z.Y., Wu, Z.L., Xing, Y.Y., Zheng, F.G., Guo, X.L., Zhang, W.G., and Hong, M.M. (1990). Nucleotide sequence of rice *waxy* gene. Nucleic Acids Res. **18**:5898.

Weber, J.A., Aldana, R., Brendan, D.G., and Edwards, J.S. (2016). Sentieon DNA pipeline for variant detection - software-only solution, over 20 faster than GATK 3.3 with identical results. Peer J Prepr **4**:e1672.

Wu, X., Liu, Y., Luo, H., Shang, L., Leng, C., Liu, Z., Li, Z., Lu, X., Cai, H., Hao, H., et al. (2022). Genomic footprints of sorghum domestication and breeding selection for multiple end uses. Mol. Plant **15**:537–551.

Wu, Y.H., Pu, X.C., Yang, C.Y., and Long, X.Y. (2010). Classification of special rice varieties and its leaf–age cultivation model in Southeast Guizhou. Guizhou Agri. Sci. **38**:49–52.

Xia, K., Ou, X., Tang, H., Wang, R., Wu, P., Jia, Y., Wei, X., Xu, X., Kang, S.H., Kim, S.K., et al. (2015). Rice microRNA osa-miR1848 targets the obtusifoliol 14α-demethylase gene OsCYP51G3 and mediates the biosynthesis of phytosterols and brassinosteroids during development and in response to stress. New Phytol. **208**:790–802.

Xu, J.S., and Li, H. (2014). Ethnic History of Greater Canton. Kunming (Yunnan People's Publishing House).

Xu, Q., Saito, H., Hirose, I., Katsura, K., Yoshitake, Y., Yokoo, T., Tsukiyama, T., Teraishi, M., Tanisaka, T., and Okumoto, Y. (2013). The effects of the photoperiod-insensitive alleles, Se13, Hd1 and Ghd7, on yield components in rice. Mol. Breed. **33**:813–819.

Xu, W., Wu, D., Yang, T., Sun, C., Wang, Z., Han, B., Wu, S., Yu, A., Chapman, M.A., Muraguri, S., et al. (2021). Genomic insights into the origin, domestication and genetic basis of agronomic traits of castor bean. Genome Biol. **22**:113.

Yan, Q.Y. (2008). Historical investigation of japonica changed to indica in southeast Guizhou Province. Ancient Modern Agricul **3**:27–34.

Yan, Q.Y. (2009). Spatial distribution and regional characteristics of flood and drought disasters in Guizhou in Ming and Qing dynasty. Agricul. History of China **28**:54–62.

Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. **88**:76–82.

Yang, W.B. (2004). From japonica to indica: a survey of grain crop cultivation and national habitat adaptation in southeast Guizhou Province. Agricul. History of China **23**:88–96.

Yang, Z.H. (2013). Sticky rice: a perspective to research ethnic minorities' history and culture in South China. Guangxi Ethnic Studies **1**:109–117.

Yang, Z.H. (2014). A historical study of the growing of glutinous rice by the Dong ethnic group: a case study in southeast Guizhou Province. J. Yunnan Minzu University **31**:53–61.

Yang, Z.H. (2021). Civilization and mutual learning: the study on glutinous rice cultural circle and its Changes. J. Minzu University of China (Philosophy and Social Sciences Edition) **48**:60–72.

Yano, M., Katayose, Y., Ashikari, M., Yamanouchi, U., Monna, L., Fuse, T., Baba, T., Yamamoto, K., Umehara, Y., Nagamura, Y., et al. (2000). Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell **12**:2473–2484.

You, X.L., and Zeng, X.S. (2010). Cultural History of Rice Cultivation in China (Shanghai: Shanghai People's Publishing House).

Yuan, H., Xu, Z., Chen, W., Deng, C., Liu, Y., Yuan, M., Gao, P., Shi, H., Tu, B., Li, T., et al. (2022). OsBSK2, a putative brassinosteroid-signaling kinase, positively controls grain size in rice. J. Exp. Bot. **73**:5529–5542.

Zafar, S.A., Patil, S.B., Uzair, M., Fang, J., Zhao, J., Guo, T., Yuan, S., Uzair, M., Luo, Q., Shi, J., et al. (2020). DEGENERATED PANICLE AND PARTIAL STERILITY 1 (DPS1) encodes a cystathionine β-synthase domain containing protein required for anther cuticle and panicle development in rice. New Phytol. **225**:356–375.

Zang, G., Zou, H., Zhang, Y., Xiang, Z., Huang, J., Luo, L., Wang, C., Lei, K., Li, X., Song, D., et al. (2016). The de-Etiolated 1 homolog of

arabidopsis modulates the ABA signaling pathway and ABA biosynthesis in rice. Plant Physiology **171**:1259–1276.

Zeng, X., Guo, Y., Xu, Q., Mascher, M., Guo, G., Li, S., Mao, L., Liu, Q., Xia, Z., Zhou, J., et al. (2018). Origin and evolution of qingke barley in Tibet. Nat. Commun. **9**:5433.

Zhang, G., Song, X., Guo, H., Wu, Y., Chen, X., and Fang, R. (2016). A small G protein as a novel component of the rice brassinosteroid signal transduction. Mol. Plant **9**:1260–1271.

Zhao, Q., Feng, Q., Lu, H., Li, Y., Wang, A., Tian, Q., Zhan, Q., Lu, Y., Zhang, L., Huang, T., et al. (2018). Pan-genome analysis highlights the extent of genomic variation in cultivated and wild rice. Nat. Genet. **50**:278–284.

Zhou, C., Zhang, Q., Chen, Y., Huang, J., Guo, Q., Li, Y., Wang, W., Qiu, Y., Guan, W., Zhang, J., et al. (2021). Balancing selection and wild gene pool contribute to resistance in global rice germplasm against planthopper. J. Integr. Plant Biol. **63**:1695–1711.

Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. **44**:821–824.

Zhou, Y., Zhao, X., Li, Y., Xu, J., Bi, A., Kang, L., Xu, D., Chen, H., Wang, Y., Wang, Y.G., et al. (2020). *Triticum* population sequencing provides insights into wheat adaptation. Nat. Genet. **52**:1412–1422.

Zhu, C.L., Xing, B., Teng, S.Z., Deng, C., Shen, Z.Y., Ai, P.F., Lu, T.G., Zhang, S.W., and Zhang, Z.G. (2021). OsRELA regulates leaf inclination by repressing the transcriptional activity of OsLIC in Rice. Front. Plant Sci. **12**:760041.

Zong, W., Ren, D., Huang, M., Sun, K., Feng, J., Zhao, J., Xiao, D., Xie, W., Liu, S., Zhang, H., et al. (2021). Strong photoperiod sensitivity is controlled by cooperation and competition among Hd1, Ghd7 and DTH8 in Rice Heading. New Phytol. **229**:1635–1649.