Plant Communications
Research Article

CellPress
Partner Journal

Genomic insights on the contribution of introgressions from Xian/*Indica* to the genetic improvement of Geng/*Japonica* rice cultivars

Di Cui[1,8], Han Zhou[2,3,8], Xiaoding Ma[1,8], Zechuan Lin[2,3], Linhua Sun[2,3], Bing Han[1], Maomao Li[4], Jianchang Sun[5], Jin Liu[4], Guixiu Jin[6], Xianju Wang[7], Guilan Cao[1], Xing Wang Deng[2,3], Hang He[2,3,*] and Longzhi Han[1,*]

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

[2]School of Advanced Agriculture Sciences and School of Life Sciences, State Key Laboratory of Protein and Plant Gene Research, Peking University, Beijing 100871, China

[3]Peking University Institute of Advanced Agricultural Sciences, Weifang, Shandong, 261325, China

[4]Rice Research Institute, Jiangxi Academy of Agricultural Sciences, Nanchang 330200, China

[5]Institute of Crop Research, Ningxia Academy of Agricultural and Forestry Sciences, Yongning 750105, China

[6]Rice Research Institute, Linyi Academy of Agricultural Sciences, Shandong Linyi 276012, China

[7]Rice Research Institute of Liaoning Province, Shenyang 110161, China

[8]These authors contributed equally to this work.

***Correspondence: Hang He (hehang@pku.edu.cn), Longzhi Han (hanlongzhi@caas.cn)**

https://doi.org/10.1016/j.xplc.2022.100325

ABSTRACT

Hybridization between Xian/*indica* (XI) and Geng/*japonica* (GJ) rice combined with utilization of plant ideotypes has greatly contributed to yield improvements in modern GJ rice in China over the past 50 years. To explore the genomic basis of improved yield and disease resistance in GJ rice, we conducted a large-scale genomic landscape analysis of 816 elite GJ cultivars representing multiple eras of germplasm from China. We detected consistently increasing introgressions from three XI subpopulations into GJ cultivars since the 1980s and found that the XI genome introgressions significantly increased the grain number per panicle (GN) and decreased the panicle number per plant. This contributed to the improvement of plant type during modern breeding, changing multi-tiller plants to moderate tiller plants with a large panicle size and increasing the blast resistance. Notably, we found that key gene haplotypes controlling plant architecture, yield components, and pest and disease resistance, including *IPA1*, *SMG1*, *DEP3*, *Pib*, *Pi-d2*, and *Bph3*, were introduced from XI rice by introgression. By GWAS analysis, we detected a GN-related gene *Gnd5*, which had been consistently introgressed from XI into GJ cultivars since the 1980s. *Gnd5* is a GRAS transcription factor gene, and *Gnd5* knockout mutants showed a significant reduction in GN. The estimated genetic effects of genes varied among different breeding locations, which explained the distinct introgression levels of XI gene haplotypes, including *Gnd5*, *DEP3*, etc., to these GJ breeding pedigrees. These findings reveal the genomic contributions of introgressions from XI to the trait improvements of GJ rice cultivars and provide new insights for future rice genomic breeding.

Keywords: Geng/*japonica* rice breeding, intersubspecific hybridization, introgression, GWAS, pedigree

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

INTRODUCTION

Asian cultivated rice (*Oryza sativa*), which can be divided into the two major subspecies Geng/*japonica* (GJ) and Xian/*indica* (XI), is one of the most important staple food crops worldwide (Xing and Zhang, 2010), supporting 21% of the total calorie intake of the global population (Fitzgerald et al., 2009). China is the largest producer of GJ rice in the world, holding the

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

largest planting area as well. In China, the planting area of GJ rice is 9.87 million square hectometers (hm^2), accounting for about 32.9% of the total rice planting area (Tang and Chen, 2021). In recent years, with increasing demand for GJ rice, there have been remarkable trends of "replace wheat with rice" in the north of China and "switch to GJ rice from XI rice" in the south of China (Zhang et al., 2013). Therefore, breeding improvement of GJ rice to keep pace with growing human needs is necessary for the expansion of GJ rice cultivation into new areas. Continued global population increases, climate and environmental changes, and a steady decrease in arable land pose enormous challenges for agriculture; new approaches for achieving genome-based breeding by design are urgently needed to develop new GJ varieties with higher productivity, better quality, stronger stress tolerance, and higher nutrient use efficiency. The rational design of breeding strategies based on genome-wide loci selection requires more understanding of the genetic basis of improvements in key agronomic traits of GJ rice cultivars.

Hybridization between XI and GJ rice combined with utilization of plant ideotypes has greatly contributed to yield improvements in modern GJ rice in China over the past 50 years (Chen et al., 2007). Owing to long-term differentiation, selection, and adaptation, both XI and GJ rice contain many favorable genes, most of which are distributed in different subspecies or ecotypes in the form of allelic variation. Therefore, combining the favorable genes of the two subspecies, including those related to yield, quality, resistance, adaptability, etc., by XI-GJ hybridization has great value for creating genotypes with greater yield potential, stronger stress resistance, and better quality (Gu, 2010). In recent years, a series of high-yielding and good-quality GJ cultivars bred in northern China have been obtained from hybridization of XI and GJ. These varieties show ideal shoot architecture with reduced tiller number, few unproductive tillers, increased panicle size, thick and sturdy culms, and improved grain yield (Bai et al., 2018). The introduction of the XI pedigree has expanded the gene pool of GJ cultivars and further increased yield, but the genomic regions and genes of the XI genome that have been introgressed into GJ cultivars in northern China are unclear.

The genomic analysis of rice germplasm resources based on resequencing is considered to be an efficient approach to address this question and has been widely used to identify genetic loci that have contributed to rice domestication (Xu et al., 2011; Huang et al., 2012; Wang et al., 2018) and improvement (Xie et al., 2015; Yano et al., 2016; Li et al., 2020). Several recent studies reported that extensive introgressions between modern GJ and XI rice varieties have been found during breeding, and introgression from GJ into XI was much greater than introgression from XI into GJ (Santos et al., 2019; Chen et al., 2020; Zhang et al., 2021). This intersubspecific introgression is likely to have increased within-population diversity and reduced subspecific differentiation (Zhang et al., 2021). Researchers also found extensive introgression loci between different subspecies or subpopulations that may have contributed to important agronomic traits; such loci include *Hd3a* and *RFT1* for heading date, *Waxy* and *ALK* for grain quality, *Pi-ta* and *Ptr* for rice blast resistance, *Xa23* for bacterial blight resistance, and *BPH29* for brown planthopper susceptibility (McNally et al., 2009; Zhao et al., 2010; Chen et al., 2020; Wei et al., 2021). However, existing knowledge about the genomic contributions of XI genome introgression to trait improvement in GJ rice is still limited.

Here, 816 GJ rice varieties were collected and sequenced, including 145 rice landraces and 671 improved varieties. The improved varieties were divided into three groups on the basis of their release date: early-stage varieties (released before 1980, hereafter B1980s), middle-stage varieties (released from 1980 to 1999, hereafter 1980s&90s), and late-stage varieties (released after 2000, hereafter A2000s). We generated a whole-genome atlas of introgression and selection during GJ breeding. Using phenotypic data collected in multiple environments, we identified loci/genes related to the improvement of agronomic traits. The novel gene *Gnd5*, which was related to grain number per panicle (GN), was identified as an introgression target from XI into GJ. The superior XI haplotypes of *Gnd5* significantly increased GN. The *gnd5* knockout mutant showed a dramatic reduction in GN, which may have been caused by a reduced level of cytokinin according to RNA sequencing (RNA-seq) analysis. We investigated the haplotype usage pattern of GN-related genes for three representative pedigrees and analyzed their genetic effects in diverse genetic backgrounds. Our work reveals the genomic contributions of introgressions from XI to trait improvements in GJ rice and provides new insights to guide molecular breeding by genome-wide design in the future.

RESULTS

Genomic analysis of GJ rice population

Our selected set of 816 GJ rice accessions was composed of 145 landraces and 671 rice improved varieties (including elite varieties, founder cultivars, and their derivatives) (Figure 1A and Supplemental Table 1). We selected 318 representative accessions and phenotyped eight key agronomic traits related to yield components, plant architecture, and environmental adaptation across six agro-ecologically diverse locations. Consistent with previous reports (Zhou et al., 2019), the most significant differences were observed in plant architecture, especially panicle traits. GN increased from 108.8 to 137.6, panicle length (PL) increased from 18.4 to 19.0 cm, and panicle number per plant (PN) decreased from 15.3 to 13.0, indicating that the panicle type had transitioned from the multi-panicle type to the large-panicle type over the course of breeding (Figure 1B). Our findings were consistent with previous reports that a series of high-yielding GJ cultivars bred in recent years in northern China showed ideal shoot architecture, with few unproductive tillers, increased panicle size, thick and sturdy culms, and improved grain yield (Bai et al., 2018). In addition, the grain length-to-width ratio (LW) increased greatly from 2.14 to 2.30. Notably, the susceptibility to panicle blast (SPB) decreased significantly (Figure 1B). Other yield component traits, including the seed setting rate (SSR) and thousand grain weight (GW), did not show significant changes during the breeding process (Supplemental Figure 1A and 1B). Moreover, plant height (PH) first decreased and then increased (Supplemental Figure 1C) because breeders were focused on generating dwarf plants during early- to middle-stage breeding, whereas they focused on increasing biomass during late-stage breeding (Gu, 2010). The trend in heading date (HD) changes

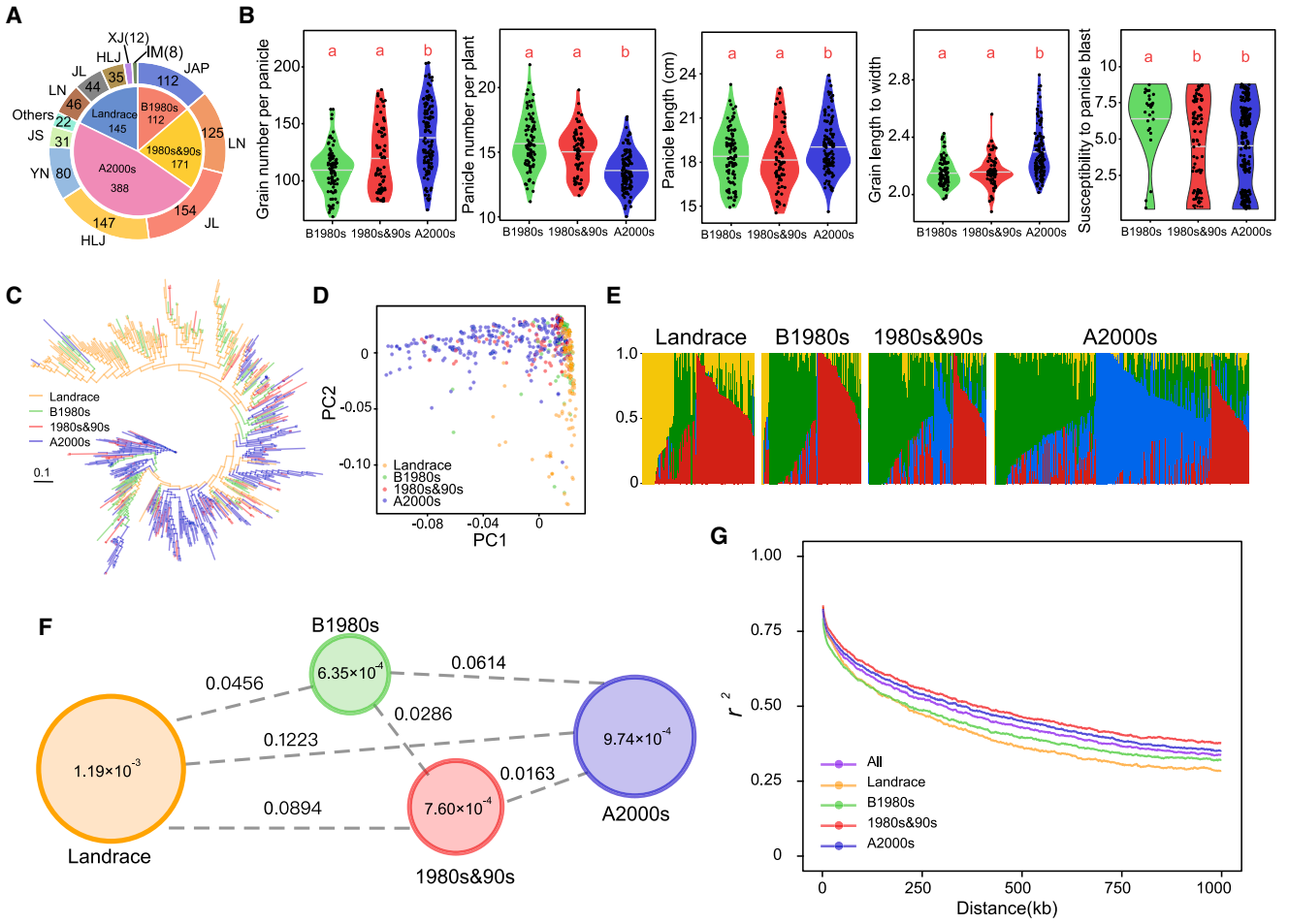

Figure 1. Phenotypic divergence, population structure, and genetic diversity of GJ rice accessions.

(A) Geographic distribution and breeding stage division of 816 resequenced GJ rice samples. Japan, JAP; Liaoning, LN; Jilin, JL; Heilongjiang, HLJ; Yunnan, YN; Jiangsu, JS; Xinjiang, XJ; Inner Mongolia, IM. B1980s, rice varieties released before 1980; 1980s&90s, rice varieties released from 1980 to 1999; A2000s, rice varieties released after 2000.

(B) Phenotypic distributions of GN, grain number per panicle; PN, panicle number per plant; PL, panicle length; LW, grain length to width ratio; and SPB, susceptibility to panicle blast (from left to right) among different breeding stages. Different letters above the boxes indicate significant differences ($p < 0.05$, Bonferroni correction) in multiple comparison testing. Violin plots show the distributions of these four agronomic traits.

(C) Maximum likelihood tree based on rice accessions consisting of landraces and improved varieties from three breeding stages.

(D) PCA plots of the first two components for rice cultivars. PC1 and PC2 indicate the scores of principal components 1 and 2, respectively.

(E) Population structure of rice cultivars (K = 4).

(F) Genetic diversity and subgroup differentiation in landraces and improved varieties from different breeding stages. The size of each circle and the number in each circle represent the level of genetic diversity (π) of each subgroup, and the number on each dashed line represents the F_{ST} value between the subgroups it connects.

(G) Genome-wide average linkage disequilibrium of rice accessions from different subgroups.

during the breeding process was similar to that in PH (Supplemental Figure 1C and 1D), consistent with the findings of a previous study (Li et al., 2020).

To characterize the genetic basis of these phenotypic changes, all 816 accessions were sequenced to an average depth of 21.2× (Supplemental Table 1), yielding 6 990 489 high-quality single-nucleotide polymorphisms (SNPs) and 1 108 456 small (<10 base pairs [bp]) indels after application of a strict filtering pipeline (Supplemental Table 2). The phylogenetic tree showed that the rice accessions B1980s tended to cluster with landraces, whereas the improved varieties, 1980s&90s and A2000s, clustered together (Figure 1C). The first two principal components (PCs) explained 25.9% of the differentiation of the rice accessions; PC1 separated the rice accessions released before the 1980s and after the 1980s, and PC2 separated landraces and cultivars to some extent (Figure 1D). Population structure analysis indicated that one of the main genetic components of rice landraces (colored in yellow) has nearly disappeared in modern cultivars, whereas genetic components with low proportions (colored in blue) have gradually expanded since the 1980s (Figure 1E, Supplemental Figures 2 and 3).

To gain a deeper understanding of the genetic makeup of the 816 rice accessions, ancestry inference analysis was performed

using the integrated 5K-rice population as a reference (Huang et al., 2010; Huang et al., 2011; Huang et al., 2012; Chen et al., 2014; Genomes project, 2014; Yano et al., 2016). The 5K-rice population includes the major *Oryza* subpopulations: wild rice; aromatic; temperate GJ; tropical GJ; intermediate GJ; XI I, II, and III; *aus*; and the mixed type. Most of the subpopulation components of our rice accessions were found to originate from temperate GJ, although components from tropical GJ and XI I, II, and III were also detected, suggesting the existence of gene flows from other subpopulations to GJ cultivars during modern breeding (Supplemental Table 3 and Supplemental Figure 4) or segregating standing variation, lineage sorting, and ancestral population structure. Compared with the genetic components from tropical GJ, the proportions of genetic components from XI I, II, and III in our rice samples were closely correlated with PC1 (XI I: R^2 = 0.12, p < 0.0001; XI II: R^2 = 0.52, p < 0.0001; and XI III: R^2 = 0.31, p < 0.0001), indicating that population differentiation between different rice breeding eras may be associated with XI genome introgression (Supplemental Figure 5).

Rice landraces show higher genetic diversity than modern varieties. A2000s varieties show greater genetic diversity (π = 9.74 × 10^{-4}) than 1980s&90s varieties (π = 7.60 × 10^{-4}) and B1980s varieties (π = 6.35 × 10^{-4}), suggesting that the introduction of germplasm from XI rice contributed to the genetic diversity of GJ rice during modern breeding (Figure 1F). Linkage disequilibrium (LD) decayed within 360 kb for rice landraces. For B1980s, 1980s&90s, and A2000s cultivars, LD decayed within 480, 680, and 640 kb, respectively (Figure 1G), in general agreement with results from a previous study (Li et al., 2020).

XI genome introgression to GJ cultivars contributes to trait improvements

The ancestral components from admixture analysis were used as a proxy for admixture proportion. The inference of admixture proportions revealed that the proportion of XI (including XI I, II, and III) genetic components was 0.069 for cultivars bred in A2000s, whereas this proportion was 0.033 for those bred in 1980s&90s and 0.017 for those bred in B1980s, consistent with the historical time period of XI-GJ hybridization (Sun et al., 2012). With regard to rice accessions from different geographic locations, the average proportions of XI genome components were ranked as follows (from maximum to minimum): Liaoning (0.100), Jilin (0.052), Heilongjiang (0.024), and Japan (0.011) (Supplemental Table 3). These results reveal that the use of intersubspecies XI-GJ hybridization varied among different locations. Four-taxon f_d statistics were calculated to detect introgression regions from XI into GJ cultivars in different breeding eras and geographic locations (Figure 2A and Supplemental Figure 6). The most introgression regions were detected from XI II into GJ rice bred in A2000s and Liaoning, and this result was further supported by TreeMix (Figure 2B and Supplemental Figure 7). For rice accessions from A2000s, we detected 45.47, 51.82, and 36.02 Mb of introgression regions from XI I, II, and III, respectively (Supplemental Table 4). In total, we found that 94.84 Mb of introgression regions could be of XI origin. For rice accessions from Liaoning, 56.51, 67.86, and 60.73 Mb of introgression regions were identified from XI I, II, and III, respectively (Supplemental Table 5).

The proportion of XI genome introgressions was significantly positively correlated with GN (XI I, r = 0.233; XI II, r = 0.489; XI III, r = 0.186; p < 0.01), PL (XI II, r = 0.222; XI III, r = 0.326; p < 0.01), and LW (XI II, r = 0.489; p < 0.01), whereas a significant negative correlation was observed between the proportion of XI genome introgressions and PN (XI I, r = −0.170; XI II, r = −0.359; XI III, r = −0.213; p < 0.01) and SPB (XI I, r = −0.186; XI II, r = −0.343; XI III, r = −0.166; p < 0.01) (Supplemental Figure 8). Introgressions from the XI genome may have contributed to improvements in yield-related traits and rice blast resistance in modern GJ cultivars during breeding. The introgression regions harbored a large number of panicle-related genes (Supplemental Tables 6 and 7). For example, the introgression region from XI I between 24.62 and 25.96 Mb on chromosome 8 contained the ideal plant architecture gene *IPA1* (Supplemental Figure 9A), which influences rice plant architecture and yield (Jiao et al., 2010). Haplotype analysis of *IPA1* based on causal variants revealed that the frequency of the superior XI haplotype Hap2 increased during GJ rice breeding (Figure 2C–2F), suggesting *IPA1* as an introgression target during modern GJ rice breeding. We also found that introgression regions from XI carried *SMG1* and *DEP3*, and the superior XI alleles (Supplemental Figures 10 and 11), which have been reported to improve yield by increasing GN (Durand et al., 2011; Qiao et al., 2011; Xu et al., 2018), increased during breeding. In addition, many other genes related to plant type and yield components, including *DLT* (Tong et al., 2012), *Ghd7* (Xue et al., 2008), and *GW5* (Liu et al., 2017), were detected in introgression regions. Notably, the frequency of the superior XI haplotypes of *Gnd7* (Hap2, conferring increased GN; ^{GN}Hap1 = 105.7 and ^{GN}Hap2 = 128.7; p < 0.01) and *GW5* (Hap2, conferring increased LW; ^{LW}Hap1 = 2.18 and ^{LW}Hap2 = 2.49; p < 0.01) increased during GJ rice breeding.

In addition, we identified two rice blast resistance genes, *Pib* and *Pi-d2*, on chromosome 2 and chromosome 6, as introgression targets (Supplemental Figure 9B and 9C) (Chen et al., 2006; Li et al., 2015; Vasudevan et al., 2016), which showed increases in the frequency of the superior XI haplotype over time (Figure 2C–2F). Other genes contributing to resistance to stress, disease, and insects were also identified as potential targets of introgression, including *OsPIN3t* (related to drought tolerance) (Zhang et al., 2012); *Pish* (Takahashi et al., 2010), *Pi5* (Cesari et al., 2013), and *Pia* (related to rice blast resistance) (Okuyama et al., 2011); *xa13* (Chu et al., 2006) and *Xa21* (related to bacterial blight resistance) (Song et al., 1995); and *Bph3* (related to brown planthopper resistance) (Kamolsukyunyong et al., 2013). Notably, genes controlling pollen fertility were also identified in XI into GJ introgression regions. For example, the superior haplotype of the hybrid incompatibility-related gene *DPL2* (Mizuta et al., 2010) was found to have been introduced from XI into GJ, suggesting the effort of developing intersubspecies hybrids.

Selection signatures in different GJ rice breeding stages

We performed a genome scan using a cross-population composite likelihood ratio (XP-CLR) approach to detect putative selected regions from different rice breeding eras: B1980s versus landrace (early-stage), 1980s&90s versus B1980s (middle-stage), and A2000s versus 1980s&90s (late-stage). A total of 1362 potential selected regions with an average size of 51.25 kb were detected

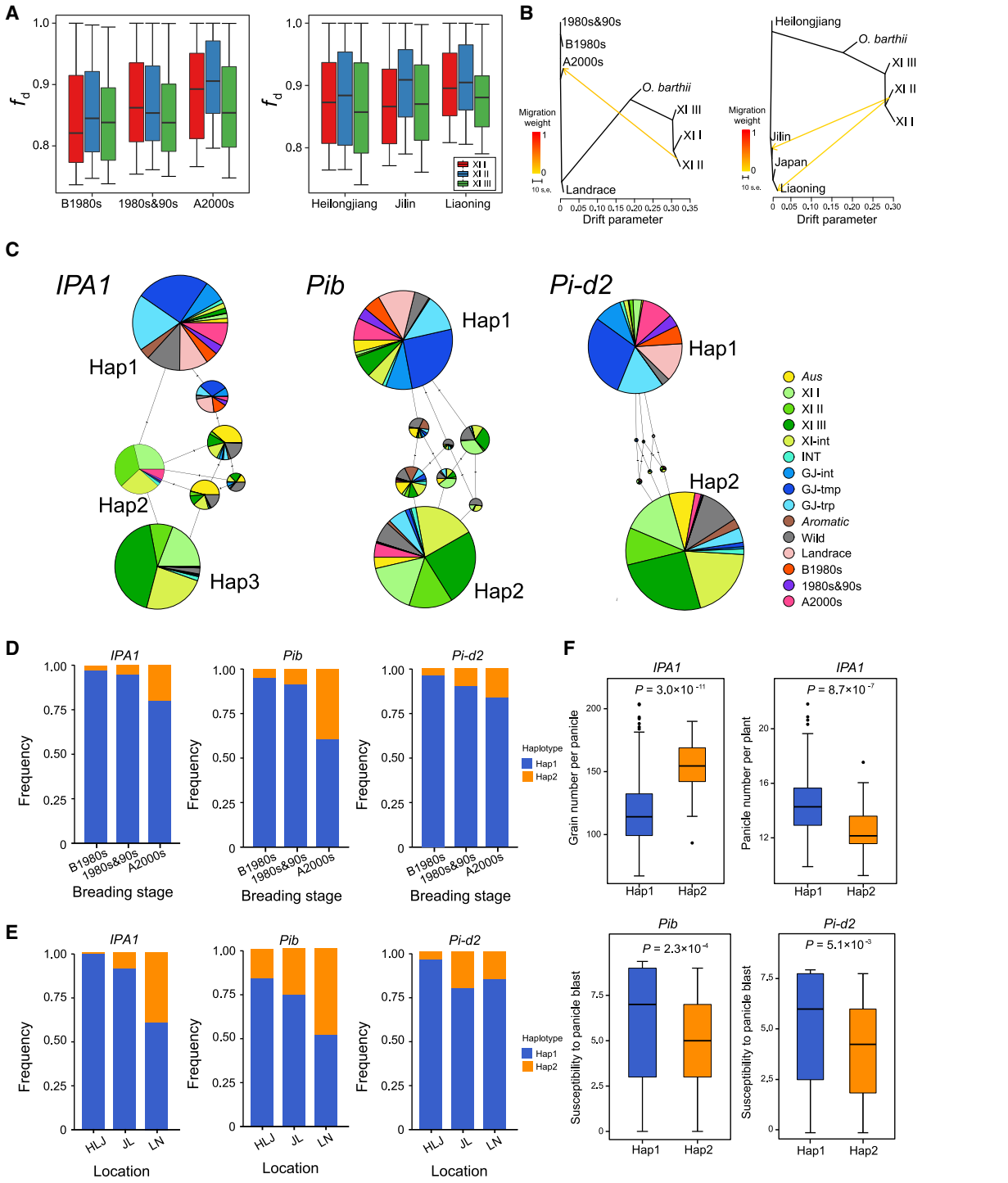

Figure 2. Identification of introgression signals from the XI genome into GJ rice cultivars in northeastern China.

(A) Distribution of the top fifth percentile f_d value in cultivars from different breeding stages (left) and different geographic locations (right). Boxplots show the median, box edges represent the first and third quartiles, and whiskers represent the 1.5× interquartile range.

(B) Maximum likelihood tree inferred by TreeMix allowing migration events from subgroups consisting of *O. barthii*, XI, and GJ samples from different breeding stages (left) and different geographic locations (right).

(C) Haplotype network of *IPA1* (left), *Pib* (middle), and *Pi-d2* (right) based on 5K-rice as a reference. The size of each circle is proportional to the number of samples for a given haplotype.

(D) Frequency distributions of *IPA1* (left), *Pib* (middle), and *Pi-d2* (right) haplotypes in different breeding stages.

(E) Frequency distributions of *IPA1* (left), *Pib* (middle), and *Pi-d2* (right) haplotypes in different locations.

(F) Boxplot of GN and PN for different haplotypes of *IPA1* (upper) and SPB for different haplotypes of *Pib* and *Pi-d2* (lower).

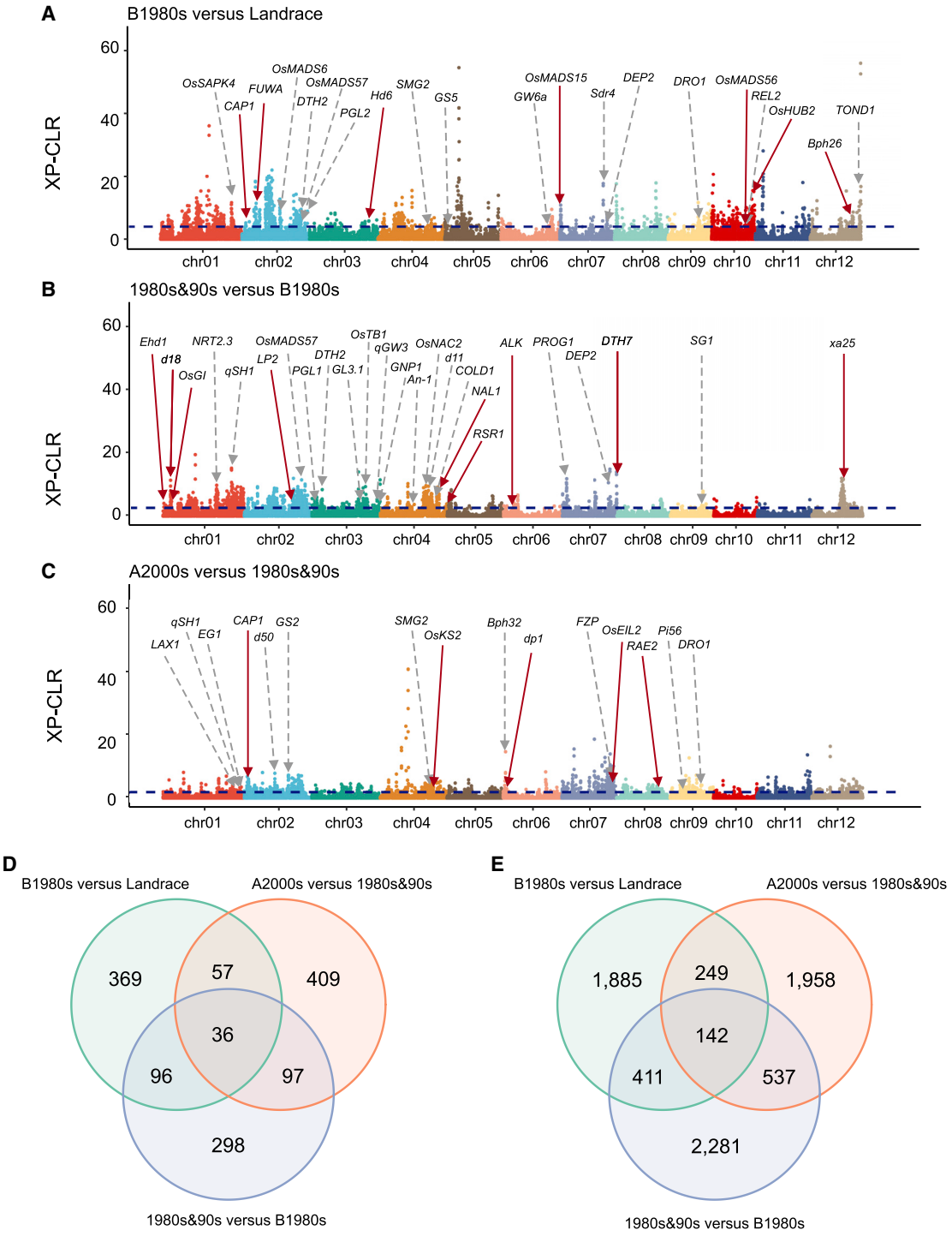

Figure 3. Whole-genome screening for differential selection during modern GJ rice breeding.

(A–C) Genome-wide selective signals (XP-CLR score) of B1980s versus landrace **(A)**, 1980s&90s verses B1980s **(B)**, and A2000s verses 1980s&90s **(C)**. The horizontal dark blue lines represent the cutoffs used to define statistical significance. The red arrows indicate candidate genes with amino acid exchange. The gray arrows with dashed lines indicate candidate genes with variants in regulatory region.

(D) Comparison of the identified selective sweeps between different breeding stages.

(E) Number of genes encompassed in the selective sweeps of different breeding stages.

in at least one of the comparisons (Figure 3A–3D and Supplemental Table 8). Because of the possibility of population structure, bottleneck, random drift, etc., we considered 286 common selective sweeps that were detected in at least two comparisons as important candidates for selection during breeding. Notably, 23.5% of our selective sweeps overlapped with previously reported sweeps for improvement in GJ rice of central China (Xiao et al., 2021), indicating that the genomic regions selected

during modern breeding are distinct among different rice cropping regions, potentially owing to environmental differences.

Across stages and populations, 7463 genes were encompassed in the selective sweeps (Figure 3E and Supplemental Table 8). Gene ontology analysis revealed that these genes were enriched in terms related to morphology, growth, and development (Supplemental Figure 12). We selected an additional subset of 1,401 genes containing amino acid exchanges in coding regions and 4329 genes with variants in regulatory regions that showed changes in allele frequency across breeding stages. Because these variations in coding regions and regulatory regions are likely to induce functional alterations, we considered these genes to be particularly intriguing candidates for selection during breeding.

As expected, many cloned genes controlling agronomically important traits were identified in selective sweeps (Supplemental Table 9). For example, *Hd6*, which influences the HD of rice plants, was located in selective sweeps in early-stage breeding. Previous studies confirmed that *Hd6* has been a major target of artificial selection during rice breeding because of its beneficial effects on flowering time and reproductive fitness during the northward expansion of rice cultivation (Nemoto et al., 2018). In addition, genes related to HD, yield, and brown planthopper resistance, including *DTH2* (Wu et al., 2013), *OsMADS57* (Guo et al., 2013), *FUWA* (Chen et al., 2015), and *Bph26* (Tamura et al., 2014), were also detected in selective sweeps. In middle-stage breeding, the pleiotropic genes *NAL1* and *DTH7* were identified in selective sweeps. As reported in previous studies, *NAL1* improved yield and altered the plant type by increasing the leaf blade width, number of secondary branches per panicle, GN, and grain yield per plant (Zhang et al., 2014), and *DTH7* greatly delayed rice heading and enhanced grain productivity (Gao et al., 2014). Notably, both of these two genes were also identified in selective sweeps for improvement in GJ rice of central China (Xiao et al., 2021). Interestingly, the domestication gene *PROG1* was identified in our selective sweeps. Previous studies reported that *PROG1* is a pleiotropic gene that can control tiller angle, GN, and PN (Tan et al., 2008). In this study, we found that the frequency of the superior haplotype (conferring increased GN) of *PROG1* had increased since the 1980s, indicating that this gene was a potential candidate of selection to improve GN during breeding. As expected, many other agronomically important genes, including *OsTB1* (Takeda et al., 2003), *DEP2* (Li et al., 2010), *ALK* (Gao et al., 2003), *RSR1* (Fu and Xue, 2010), *COLD1* (Ma et al., 2015), and *Xa25* (Liu et al., 2011), were also included in the selective sweeps. In late-stage breeding, key genes related to yield, disease resistance, and insect resistance were identified as potential targets of selection, including *LAX1* (Komatsu et al., 2001), *FZP* (Bai et al., 2017), *RAE2* (Jin et al., 2016), *GS2* (Hu et al., 2015b), *SMG2* (Xu et al., 2018), *Pi56* (Liu et al., 2013), and *Bph32* (Ren et al., 2016). The frequency of the superior haplotype of each gene increased during breeding, indicating that these genes had undergone breeding pressure (Supplemental Figure 13). Intriguingly, we found that some genes (including *DTH2*, *DEP2*, *SMG2*, etc.) in selective sweep regions were detected in at least two breeding stages and may therefore have undergone consistent selection during GJ breeding.

Genome-wide association study of agronomic traits in GJ rice

We performed a genome-wide association study (GWAS) of eight key agronomic traits using 318 representative accessions from all breeding eras and identified a total of 250 significant loci ($p < 1 \times 10^{-4}$; Supplemental Figure 14 and Supplemental Table 10). Among them, 100 significant loci overlapped with known functional genes or previously reported quantitative trait nucleotides (QTNs) (Supplemental Table 10). For example, *NAL1*, *DEP1*, *Ghd7*, and *GS3* (Fan et al., 2006) were successfully identified in these GWAS peaks (Supplemental Table 11). We also identified some new and highly promising associations, such as *OsFTL6* (Os04g0488400), which encodes a protein homologous to *Arabidopsis* FT (flowering time) and was associated with HD. In addition, we found that *OsGA2* (Os02g0571300), encoding a protein homologous to *Arabidopsis GA2*, which functions in the gibberellin biosynthetic pathway in a manner related to regulation of PH, was associated with PH. For each of these two loci, putative causal polymorphisms were also identified in their coding regions. By combining the introgression and selection analysis mentioned above, we found that *OsFTL6* has been a potential selection target of GJ breeding, and *OsGA2* showed evidence of introgression during the breeding process (Supplemental Table 10).

The intersections of regions subject to introgression or selection with GWAS peaks revealed loci related to trait improvement in modern breeding. In total, 107 and 146 of the GWAS loci overlapped with genomic regions showing evidence of introgression and selection, respectively. Among the GWAS loci associated with panicle traits, 41.2% (21/51) and 62.7% (32/51) (23 for PL, 17 for PN, and 11 for GN; Supplemental Table 10) showed evidence of introgression and selection, respectively, which were important for modern breeding. For example, a series of genes with known functions were successfully identified in both GWAS peaks and introgression regions, including *OsLIC1* (Wang et al., 2008) and *PHS9* (Garg et al., 2010) for GN. Haplotype analysis further revealed that the frequency of the superior XI haplotype Hap2 of these genes increased during GJ rice breeding (Supplemental Figure 15), suggesting that these genes were introgression targets during modern GJ rice breeding. In addition, *DTH7* and *NAL1*, which were related to PH and PN, were identified in both GWAS peaks and selective sweeps, and the frequency of their superior haplotypes increased during GJ rice breeding (Supplemental Figure 13).

A novel GRAS transcription factor gene was identified as an introgression target

We observed an SNP (GN_Chr05_Pos18409450; $p < 1 \times 10^{-4}$) with a significant signal associated with GN on chromosome 5 from 18 241 644 to 18 409 696 within a large LD block (>300 kb) (Figure 4A and 4B) that also showed strong evidence of introgression by f_d analysis in the middle stages of GJ breeding (Supplemental Table 4). We further narrowed our focus to nine genes with nonsynonymous variants in this block. A novel GRAS transcription factor gene named *Gnd5* (Os05g0378500) was identified as a high-confidence candidate.

Previous studies have demonstrated that proteins in the GRAS family are involved in gibberellin signal transduction, which plays

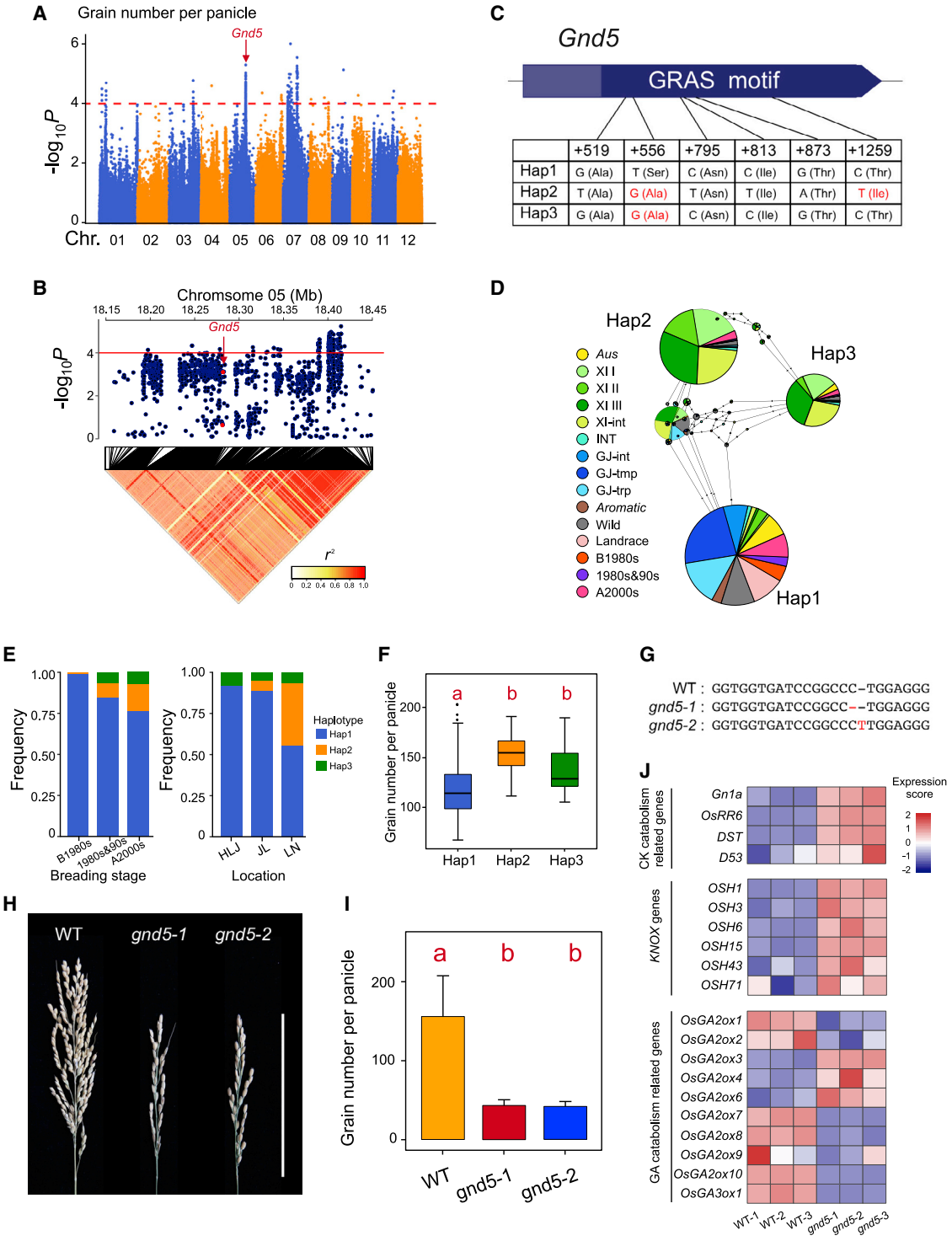

Figure 4. GWAS for grain number per panicle and identification of the novel gene *Gnd5* on chromosome 5.

(A) Manhattan plot for GWAS results for grain number per panicle. The red arrow indicates the location of the candidate gene *Gnd5*. The red dashed line represents the significance threshold ($-\log_{10}P = 4$).

(B) Local Manhattan plot (top) and LD heatmap (bottom) surrounding the peak on chromosome 5. Red dots indicate the position of missense variations in *Gnd5*.

(C) Gene structure of *Gnd5* and DNA polymorphisms in the coding region.

(D) Haplotype network of *Gnd5* based on 5K-rice as a reference.

(E) Frequency distributions of *Gnd5* haplotypes in different breeding stages (left) and different locations (right).

(F) Boxplot of GN for different haplotypes of *Gnd5*.

(legend continued on next page)

critical roles in many biological processes during plant development (Tian et al., 2004). *Gnd5* encodes a GRAS protein that belongs to the SCL3 subfamily (Supplemental Figure 16) (Niu et al., 2019) and contains a 1683-bp open reading frame (ORF) with only one exon. Among our 816 rice accessions, only two nonsynonymous SNP variants were detected in the coding region (Figure 4C). Haplotype analysis based on these two SNPs revealed that the frequency of the superior haplotype (Hap2, conferring increased GN) increased during GJ rice breeding. Notably, this superior haplotype was introduced from XI rice into GJ rice in the 1980s, indicating that *Gnd5* has been an introgression target during modern GJ rice breeding (Figure 4D–4F). To validate the effect of *Gnd5* on GN, we assayed phenotypes using CRISPR-Cas9 knockout lines that contained homozygous indels in exons (*gnd5-1* and *gnd5-2* had a 1 bp in-frame deletion or insertion at +627 bp of the *Gnd5* locus, respectively). GN was dramatically reduced in *gnd5-1* and *gnd5-2* knockout mutants by 72.21% and 73.23% ($p < 0.01$; Figure 4G–4I), demonstrating a critical role for *Gnd5* in the regulation of GN.

We also performed RNA-seq analysis on young panicles from *gnd5* mutants and wild-type plants and identified 2975 putative differentially expressed genes (DEGs), 1625 upregulated and 1350 downregulated (Supplemental Figures 17 and 18, Supplemental Table 12). Gene ontology analysis revealed that these DEGs were highly enriched in the term response to hormone ($p = 5.2 \times 10^{-11}$), and most of the hormone response genes were in the cytokinin signaling pathway, suggesting that *gnd5* was probably involved in the cytokinin signaling pathway (Supplemental Figure 19). In the *gnd5* mutants, we observed increased expression of cytokinin-catabolism-related genes such as *Gn1a*, *OsRR6*, *DST*, and *D53* (Figure 4J), which led to a reduction of cytokinins in reproductive meristem activity, resulting in inhibition of cell division and reduced GN. As reported previously, cytokinins and gibberellins play antagonistic roles in the regulation of reproductive meristem activity, and the balance between these two hormones is regulated by KNOX proteins (Wu et al., 2016). We found increased expression levels of *KNOX* genes (*OSH1*, *OSH3*, *OSH6*, *OSH15*, *OSH43*, and *OSH71*) and decreased expression levels of gibberellin-catabolism-related genes (*OsGA2ox1*, *OsGA2ox2*, *OsGA2ox7*, *OsGA2ox8*, *OsGA2ox9*, *OsGA2ox10*, and *OsGA3ox1*) in *gnd5* mutants, consistent with the reduced-GN phenotype, suggesting that *Gnd5* plays roles in gibberellin signaling.

Recent structural studies have shown that the conserved GRAS domain mediates homodimerization (Hakoshima 2018). First, we used co-immunoprecipitation (co-IP) and bimolecular fluorescence complementation (BiFC) assays to confirm that OsGnd5 showed homomeric interactions in the nucleus and cytosol (Supplemental Figure 20A and 20B). Notably, one of the amino acid substitutions (S185A) was observed in Hap2 and Hap3 in the Leucine heptad repeat I(LHRI) domains, and the other amino acid substitution (T420I) was detected in Hap2 in the PFYRE domains. Intriguingly, both the single amino acid substitution (S185A) and the double amino acid substitution (S185A and T420I) impaired the homomeric interaction of OsGnd5, and this result was confirmed by yeast two-hybrid and co-IP assays (Supplemental Figure 20C and 20D). To further examine the impact of different haplotype versions of *Gnd5* on gene expression regulation of panicles, we transiently expressed OsGnd5, OsGnd5^{S185A}, and OsGnd5$^{S185A+T420I}$ and examined whether they had distinct activities in the regulation of *Gn1a*, *OsRR6*, and *DST*. As shown in Supplemental Figure 20E, OsGnd5^{S185A} and OsGnd5$^{S185A+T420I}$ showed higher activity in repressing the expression of *Gn1a*, *OsRR6*, and *DST*. Taken together, these results indicate that the amino acid substitution in Hap2 and Hap3 in the LHRI and PFYRE domains reduced the homomeric interaction of OsGnd5, further improving its repression activity towards genes related to cytokinin catabolism and increasing GN.

Superior haplotype combinations were used differently in three GJ breeding pedigrees

To explore the genetic factors that have affected trait improvement in GJ rice breeding, we examined the haplotype usage patterns of genes related to GN in three historical and representative pedigrees of founder cultivars, including Liaogeng5 (LG5), Jigeng60 (JG60), and Hejiang20 (HJ20), which originated from Liaoning, Jilin, and Heilongjiang, respectively (Figure 5A–5C). The haplotype usage of 32 known functional genes related to GN and the newly identified gene *Gnd5* was calculated under multiple environments. For most of the 33 genes, several superior haplotypes, such as those of *NAL1, Hd1, ASP1, GNP1, Ghd7,* and *Ghd8,* were widely used in all three pedigrees. In particular, the superior haplotypes of *Hd1* and *Ghd8* were found to be nearly fixed; however, some haplotypes were used differently among the three selected pedigrees. For example, the superior haplotypes of *Gnd5*, *IPA1*, *PAP2*, and *SP1* were widely identified in the LG5 pedigree, but they were present in only a small number of accessions in the JG60 and HJ20 pedigrees. The superior haplotypes of *Gn1a* and *LAX1*, which are associated with increased GN, were found in most accessions in the JG60 and HJ20 pedigrees. Many accessions in the LG5 pedigree, such as Liaogeng5 and Shennong189, lacked the superior haplotypes of *Gn1a and LAX1*. Notably, the founder cultivar Jigeng60 and super rice cultivars, such as Liaoxing1 and Jigeng88, are excellent donors for superior haplotypes of *Gn1a* and *LAX1*, which improve GN.

The genetic effects of the 33 genes were calculated in the GJ cultivars across three pedigrees and three geographic locations, including Liaoning, Jilin, and Heilongjiang (Figure 5D and Supplemental Figure 21A). In the LG5 pedigrees, large-effect genes, such as *NAL1*, *GAD1*, *OsLSK1*, and *Gnd5*, accounted for 10.14%, 10.08%, 8.12%, and 7.61% of the phenotypic variation, respectively. *OsLSK1*, *APO1*, *DEP3*, and *OsMADS50* were large-effect genes in the JG60 pedigree, whereas *OsMADS50*, *GSN1*, and *OsSHI1* were large-effect genes in the HJ20 pedigree. In addition, some large-effect genes, such as *sped1-D*, *Gnd5*,

(G) Knockout sequences of *Gnd5* targeted by the CRISPR-Cas9 system.
(H) Image of panicles of wild-type plants (WT) and knock-out mutants (*gnd5-1* and *gnd5-2*). Scale bar, 15.5 cm.
(I) Bar plot of GN for WT and knock-out mutants (*gnd5-1* and *gnd5-2*).
(J) The heatmap of scaled expression scores of GA- and CK-related genes in the WT and the *gnd5* mutant.

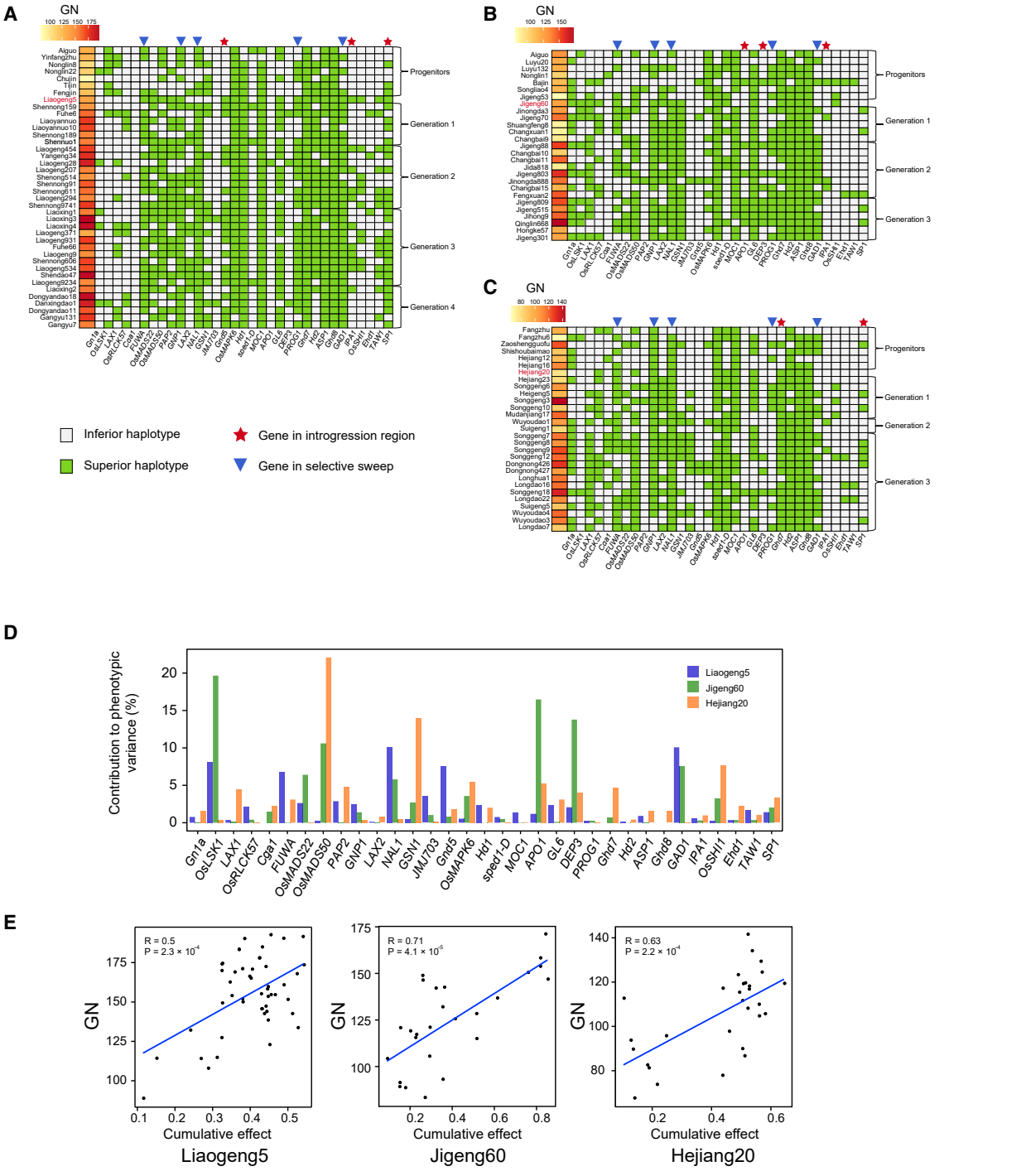

Figure 5. Usage pattern of 33 grain-number-related genes in three pedigrees and gene effects in different pedigrees.

(A–C) The phenotypes and haplotypes of 33 genes in rice accessions from the Liaogeng5 **(A)**, Jigeng60 **(B),** and Hejiang20 **(C)** pedigrees. Generations one to five represent the first to fifth filial generations of the founder cultivar, Liaogeng5, Jigeng60, or Hejiang20. Haplotypes are represented by different colors. Genes harbored in the introgression region and selective sweeps are pointed out by red stars and blue triangles, respectively.

(D) Bar plot of the contribution to phenotypic variance for each gene from different pedigrees.

(E) The correlation between the cumulative effect of 33 grain-number-related genes and the GN in three pedigrees (from left to right: Liaogeng5, Jigeng60, and Hejiang20).

and *ASP1*, accounted for 12.24%, 11.22%, and 7.10% of the phenotypic variation, respectively, in Liaoning. *APO1*, *DEP3*, and *OsMADS50* were identified as large-effect genes in Jilin, and *OsMADS50*, *PAP2*, and *Gnd5* were identified as large-effect genes in Heilongjiang. Haplotype usage patterns were consistent with the effects of genes such as *NAL1*, *APO1*, and *OsMADS50*. Notably, the superior haplotypes of *Gnd5* and *DEP3* were only specifically introgressed into LG5 and JG60

pedigrees, respectively. Therefore, estimating genetic effects in different breeding locations can explain the distinct introgression levels of XI gene haplotypes in the three GJ pedigrees. Notably, the cumulative effect of the 33 genes was significantly positively correlated with GN in three pedigrees and two locations (Figure 5E and Supplemental Figure 21B). In future breeding programs, genetic insights from existing pedigrees may enable quantitative trait loci (QTL) pyramiding and breeding route optimization in a wide rice genetic background.

DISCUSSION

Intensive selection during breeding often reduces the genetic diversity of crops, which may result in artificial genetic bottlenecks and the discarding of many favorable genetic variations (Hyten et al., 2006). Interestingly, our study showed increasing nucleotide diversity in GJ cultivars during modern breeding, in contrast to a previous study that reported reduced nucleotide diversity in maize inbred lines during modern breeding (Hufford et al., 2012) but consistent with a report of increased diversity during Chinese wheat breeding (Hao et al., 2020). The reason for the increased nucleotide diversity of modern GJ cultivars is the wide use of XI-GJ hybridization after the 1980s, which introduced XI haplotypes that were absent in GJ landraces and early-stage varieties. It is well established that the XI and GJ genomes represent quite distinct gene pools owing to long-term ecological adaptation (Gu, 2010), and, accordingly, XI and GJ rice varieties show obvious differences in physiological characteristics. Artificial crosses generated by breeding programs have led to the introgression of XI haplotypes into GJ cultivars and promoted the recombination of these two gene pools, creating new recombinant blocks and improving the agronomic traits of GJ cultivars. Hence, the genetic diversity of GJ cultivars was enhanced during modern late-stage breeding. This finding is consistent with previous research showing that intersubspecific introgression was likely to have increased within-population diversity (Zhang et al., 2021). Furthermore, we found clear population differentiation between varieties bred before the 1980s and varieties released after the 1980s, and this differentiation was caused mainly by XI genome introgression. Our findings imply that XI-GJ hybridization has made great contributions to the enhancement of genetic diversity and population differentiation in GJ cultivars.

Using a genome-wide scanning method, we found that modern breeding has increased introgression from three XI subpopulations into GJ cultivars. Notably, the proportion of XI genome introgressions was significantly positively correlated with GN and rice blast resistance, providing insight into the contribution of the XI genome to the breeding of GJ cultivars. In the early, middle, and late stages of breeding, the average genetic effects of introgression for yield components (including GN, PN, and GW) were 43.79%, 50.81%, and 56.59%, respectively (Supplemental Table 13). These results suggested that the contribution of XI genome introgression to the genetic improvement of GJ cultivars was consistently increasing during modern breeding. In the early breeding stage of GJ rice, trait improvement was achieved mainly through intrasubspecies hybridization and artificial selection to pyramid and fix the superior alleles of key genes, and the genetic background became narrower at the same time. In the mid-to-late breeding stage, to broaden the genetic basis of GJ cultivars and break the breeding bottleneck, superior alleles from the XI rice genome were introduced into GJ cultivars by crosses between XI and GJ subspecies and intensive artificial selection. With continuous optimization of trait combinations during the breeding process, superior haplotypes in both XI and GJ genomes were aggregated, and a number of new varieties with high yield, good quality, and strong resistance to diseases and pests were cultivated.

Since XI-GJ hybridization and ideal plant type breeding strategies have been widely exploited in modern rice breeding, superior haplotypes of genes from XI rice, especially those related to plant architecture and yield, have been introduced into GJ rice and have played important roles in improving modern GJ cultivars (Xing and Zhang, 2010). As expected, many haplotypes of key genes associated with plant type and yield, including *IPA1*, *DEP3*, *SMG1*, *Ghd7*, and *GW5*, were identified as potential targets of introgression, and the frequency of the superior XI haplotype for each gene increased during GJ rice breeding. Moreover, a number of stress, disease, and insect resistance controlling genes, including *OsPIN3t, Pib, Pi-d2, Pia, xa13,* and *Bph3,* were also identified as potential targets of introgression. These findings were consistent with a previous study in which genes related to rice blast resistance (*Pi-ta* and *Ptr*) and bacterial blight resistance (*Xa23*) were found in regions that had undergone introgression from XI into GJ cultivars (Chen et al., 2020). Although some superior haplotypes from XI have successfully improved the plant architecture and yield of GJ cultivars, others are underused or even unused. For example, the superior haplotypes of *IPA1* from XI were abundant in accessions from Liaoning but scarce in accessions from Heilongjiang. The same situation was observed for *Gnd5*, which has huge potential for future breeding utilization. Superior XI haplotypes of genes related to efficient fertilizer utilization, abiotic stress, and biotic stress were also unutilized. For example, the superior XI haplotypes of the nitrate transporter gene *NRT1.1B* (Hu et al., 2015a) and the nitrate reductase gene *OsNR2* (Gao et al., 2019) could greatly promote nitrogen use efficiency, but the inferior haplotypes of these two genes were almost fixed in GJ cultivars. In addition, the superior XI haplotypes of *TOND1* (providing increased tolerance to nutrient deficiency) (Zhang et al., 2015), *OsLG3* (providing increased tolerance to drought) (Yu et al., 2017), and *LHCB5* (providing increased blast resistance) (Liu et al., 2019) were also underutilized in GJ breeding. Our results suggest that many superior haplotypes of key genes related to fertilizer use efficiency as well as abiotic and biotic stress traits are currently underutilized in GJ rice, revealing the huge potential of intersubspecific hybridization for genetic improvement of GJ rice in the future. Our study not only sheds light on the genomic contributions of introgressions from XI to trait improvements of GJ rice cultivars but also provides useful information for maximizing the level of XI-GJ hybridization in crop breeding.

METHODS

Plant materials and genome resequencing

A total of 816 GJ rice varieties were collected from China National Crop Genebank, including 145 rice landraces and 671 improved varieties (consisting of elite varieties, founder cultivars, and their derivatives). The varieties were identified as GJ type both based on phenotype using Cheng's index method (Wang et al., 1987) and genotype using whole-genome

SNPs. Most of the accessions were collected in northeastern China and Japan. Detailed information about each accession, including name, geographic origin, release time, and subpopulation classification, can be found in Supplemental Table 1.

We performed DNA isolation and genome resequencing of each rice accession. The young seedlings of 816 rice accessions were collected, and their genomic DNA was extracted with the cetyl trimethylammonium bromide (CTAB) method (Murray and Thompson, 1980). The genome sequencing libraries (Illumina TruSeq DNA) were sequenced with an Illumina X Ten sequencer, yielding a total of 43.2 billion 150-bp paired-end reads (6.5 tera-bases; Supplemental Table 1). Sequencing coverage and depth for each sample were calculated with mosdepth (Pedersen and Quinlan, 2018); the average sequencing coverage for each sample was 90%, and the average depth was 21.2× (Supplemental Table 1).

The integrated 5K-rice accession was gathered from six previously published datasets: (1) a global collection of 533 diverse *O. sativa* accessions generated by Chen et al. and downloaded from the NCBI Sequence Read Archive under accession number PRJNA171289 (Chen et al., 2014); (2) 1083 worldwide rice varieties collected by Huang et al. and downloaded from the European Bioinformatics Institute (EBI) European Nucleotide Archive with accession numbers ERP000106 and ERP000729 (Huang et al., 2011); (3) 446 diverse *O. rufipogon* accessions from Asia and Oceania sequenced by Huang et al. and downloaded from the EBI European Nucleotide Archive with accession numbers ERP001143(Huang et al., 2012); (4) the 3000 rice genome project, composed of 2466 accessions from the International Rice Genebank Collection (IRGC) at the International Rice Research Institute (IRRI) and 534 accessions from the China National Crop Genebank (CNCGB), downloaded from the NCBI Sequence Read Archive under accession number PRJEB6180 (Genomes project, 2014); (5) 176 GJ rice varieties cultivated in Japan, collected by Yano et al., and downloaded from the NCBI Sequence Read Archive under accession number PRJDB4518 (Yano et al., 2016); and (6) 87 *Oryza barthii* samples from the NCBI Sequence Read Archive under accession number PRJEB21312 (Cubry et al., 2018).

Phenotyping

Three hundred and eighteen rice varieties were planted and phenotyped across six different environments in 2017: the city of Shenyang in Liaoning province, Beijing, Yinchuan in Ningxia province, Linyi in Shandong province, Nanchang in Jiangxi province, and Sanya in Hainan province in China. We collected the seeds from the same panicle of each rice variety for phenotyping and sequencing. A sequential experimental design was used in all six trials with two replications. Each plot consisted of three rows with 16 plants; each was planted with a 13.3-/26.7-cm spacing between the plants and the rows. At least five plants in the middle of the plot were selected for phenotyping of PH, PL, GN, PN, GW, SSR, and LW according to the standard evaluation system for rice (Han et al., 2006).

A total of 501 rice varieties were screened for panicle blast resistance under natural conditions in the blast nursery at the experimental farm of Donggang Agricultural Center in Liaoning province, China. The panicle blast screening was conducted twice during the two wet seasons of 2017 and 2018 with two replications. Each plot contained three rows with 15 plants per row. The row and column spacings were set to 30.0 and 13.3 cm, respectively. In addition, known susceptible checks (Menggudao) were sown in borderlines as spreader rows and after each test variety to promote uniform spread of the disease. Reactions of each variety for panicle blast were scored according to the standard evaluation system for rice (Han et al., 2006).

Read alignment and variant calling

fastp was used to trim adapters and filter low-quality reads (Chen et al., 2018) with default settings except for a minimum read length of >70 bp. The clean paired-end reads were aligned to the rice reference genome (Nipponbare reference no. IRGSP-1.0) (Kawahara et al., 2013) using default parameters of BWA (v.0.7.12) with the MEM algorithm (Li and Durbin, 2009). Reads with multiple hits were removed using SAMtools (v.1.3) (Li et al., 2009), and only the uniquely mapped reads were retained for variant detection. Potential PCR duplicates were marked using Picard (v.2.1.1) with MarkDuplicates and then removed with SAMtools. Alignments around small indels were remapped using genomic analysis toolkit (GATK) (v.3.7) (McKenna et al., 2010) with RealignerTargetCreator and IndelRealigner, and raw variants were called based on the realigned BAM file. Consecutive steps using GATK 'HaplotypeCaller' with parameters '-stand_call_conf 30, -stand_emit_conf 10', SAMtools (v.1.3) (Li et al., 2009), and BCFtools (v.1.8) (Li, 2011) were also applied to each rice accession for variant calling. To reduce the false discovery rate of the variants, the variants were filtered using GATK VariantFiltration according to the following threshold: QUAL <30, QD < 2, ReadPosRankSum < -8, FS > 60. To correct base quality scores for systematic technical errors based on the concordance variants detected by SAMtools and GATK, the variants were further filtered with GATK 'BaseRecalibrator'. After the filtrations, all of the GVCF files from 816 resequenced rice accessions were then merged. A total of 6 990 489 SNPs and 1 108 456 indels (small insertions and deletions <10 bp) were identified from the analyses of the genomes of 816 accessions. The variants with missing rate more than 0.2, minor allele frequency less than 0.05, and heterozygous genotypes were excluded using VCFtools (v.0.1.13) (Danecek et al., 2011). SNP annotation was performed using SnpEff (v.4.3) (Cingolani et al., 2012) on the basis of the Nipponbare reference genome (v.7.0). The untyped genotype data of the SNP sets used in the following analysis were imputed with Beagle software (v.4.1) (Browning and Browning, 2007) using a sliding window of 10 000 bp and a step length of 1000 bp.

Population genetic analysis

The PC analysis was estimated using PLINK (v.1.90) (Purcell et al., 2007) pca. The genome-wide average r^2 between two SNPs within 1-Mbp windows was calculated using PLINK with the parameters –ld-window 99999 –ld-window-kb 1000 –ld-window-r2 0, and the distance of LD decay was represented as the physical distance over which r^2 dropped to half of the maximum value. The sequence diversity statistics (π) and the population-differentiation statistics (F_{ST}) were computed using a 10-kb window with VCFtools (v0.1.13) (Danecek et al., 2011). An individual-based maximum likelihood tree was constructed using the GTRGAMMA model in RAxML software (v.8.2.X) (Stamatakis, 2014) with a subset of 124 299 SNPs. This SNP subset was filtered by LD pruning with a window size of 10 kb, a window step of one SNP, and an r^2 threshold of 0.8. The model-based estimation of ancestry for the population was carried out using ADMIXTURE (Zhou et al., 2011) with the evenly distributed SNPs with default parameters.

Identification of introgression

A four-taxon f_d statistic was used to identify the genomic segments introgressed from the XI group (Durand et al., 2011). *Oryza barthii* (O) was used to infer the ancestral (A) and derived (B) SNP allelic states in the populations of XI (P3), GJ landrace (P1), and modern GJ cultivars (P2). Inference of ancestral state is described above. Three previously identified XI subpopulations were used as P3: XI I, XI II, and XI III. Without gene flow, the ABBA and BABA allele configurations in the four-taxa tree (((P1, P2), P3), O), should be equally frequent; gene flow between GJ cultivars and XI rice would result in an excess of ABBA relative to BABA that could be detected using the f_d statistic. The f_d statistic was calculated in sliding windows of 100 000 kb with a step of 10 kb. Windows with fewer than three informative SNPs (neither ABBA nor BABA) were ignored. Windows with negative values of Patterson's D statistic (closely related to the f_d statistic) and f_d > 1 were ignored. TreeMix (v.1.12) (Pickrell and Pritchard, 2012) was used to build the ancestry graphs assuming one migration edge, the placement and weight of each being optimized by

the algorithm. TreeMix was run using the global option, and the sample size correction was disabled. SEs were estimated in each block with 500 SNPs. The optimal number of migration edges on population trees from TreeMix was estimated with the R package OptM (Fitak, 2021). The haplotype network of candidate genes was constructed with PopART (v.1.7) (Leigh et al., 2015) using the missense and nonsense SNPs from the coding regions (if there was no missense or nonsense SNP, SNPs within 5 kb upstream of genes from the regulatory regions were used) (Supplemental Table 14).

Detection of selective sweeps

Whole-genome screening of selection was performed using XP-CLR (v.1.0), a method based on modeling the likelihood of multi-locus allele frequency differentiation between two populations (Chen et al., 2010). We used the earlier breeding era as a reference and the later one as a query to identify the potential breeding sweeps (B1980s versus landrace, 1980s&90s versus B1980s, and A2000s versus 1980s&90s). The program XP-CLR was run with parameters "-w1 0.0005 100 100 1 -p1 0.7" for each chromosome. The XP-CLR score was calculated for each 100 bp, and the average XP-CLR score was obtained for each 10-kb segment. Adjacent segments with an average of XP-CLR scores greater than the 80th percentile of the genome-wide average XP-CLR were then grouped as putatively selected regions. We then merged putatively selected regions with distance less than 10 kb, and regions in the highest 20th percentile of these scores were considered to be selective sweeps. We combined the SNP annotations from SNPeff (Cingolani et al., 2012) and gene functional annotations to select an additional subset of 1401 genes containing amino acid exchanges in coding regions, as well as 4329 genes with variants in regulatory regions (5 kb upstream of the gene).

GO analysis

Genes harbored in the selection regions from different breeding stages were submitted to agriGO v.2.0 for gene ontology (GO) enrichment analysis with default settings (Tian et al., 2017). Fisher's exact test was used to identify potentially significantly enriched GO terms ($p < 0.05$).

GWAS of key agronomic traits

A total of 1 204 026 SNPs (minor allele frequency [MAF] $\geq$0.05, missing rate $\leq$0.2) in 318 rice accessions with phenotype data were used to carry out GWAS. The missing genotypes were imputed using Beagle (v.4.1) (Browning and Browning, 2007) with default parameters. We performed GWAS for eight agronomic traits, GN, GW, HD, PH, PN, PL, LW, and SSR, using the best linear unbiased prediction (BLUP) values (Bates et al., 2014). The GWAS was conducted with a linear mixed model that was implemented in the EMMAX package (Kang et al., 2010). The top three PCs were used as covariates. We selected 1×10^{-4} (Benjamini–Hochberg false discovery rate [FDR] <0.05) as the genome-wide significance cutoff. We then enlarged the candidate region to 100 kb centered on the GWAS signal peak to identify candidate genes.

Phylogenetic analysis of GRAS genes

The GRAS gene list was downloaded from the Plant Transcription Factor Database (PlantTFDB v5.0). The sequences of GRAS proteins from *Arabidopsis* and rice were downloaded from the Arabidopsis Information Resource (https://www.arabidopsis.org/) and the Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/index.shtml), respectively. Multiple alignments and neighbor-joining tree construction were performed with MEGA software (v.7.0) (Kumar et al., 2016).

Targeted mutagenesis using the CRISPR-Cas9 system

The mutant was generated using CRISPR-Cas9 technology in Zhonghua 11. In brief, gene-specific single-guide RNAs (sgRNAs) were designed online at http://skl.scau.edu.cn/targetdesign/, and each sgRNA cassette was separately cloned into *pYLCRISPR/Cas9Pubi-H*. The resulting constructs were separately introduced into wild-type plants via *Agrobacterium*-mediated transformation. The resulting plants were selected in Yoshida's culture solution supplemented with 50 mg L^{-1} hygromycin, and all mutations were confirmed by Sanger sequencing. T_1 seeds lacking hygromycin resistance were used for subsequent experiments.

RNA-seq data analysis

We collected three biological replicate samples of 1-cm young panicles from WT and *gnd5*, and RNA was extracted following the ZR Plant RNA Miniprep (ZYMO). The samples were considered to be of adequate concentration and quality when the value of optical density $(OD)_{260/280}$ was greater than 1.9 and less than 2.1, with a concentration $\geq$200 ng/μl. The quality-checked RNA was used to construct an RNA (polyA-selected) Illumina library and sequenced using the HiSeq 2500 high-throughput sequencing platform (Illumina, San Diego, CA, USA). The RNA-seq data from the WT and *gnd5* were processed with fastp (v.0.20.0) (Chen et al., 2018) to trim adapters and filter low-quality reads with default settings, except for a minimum read length of >70 bp. The clean paired-end reads were aligned to the rice reference genome (Nipponbare reference no. IRGSP-1.0) using STAR (v.2.7.9a) (Dobin et al., 2013). featureCounts was used to obtain the mapped read counts for each gene (Liao et al., 2014). DESeq2 (v.1.24.0) (Love et al., 2014) was used to identify the DEGs.

Co-immunoprecipitation (co-IP) assay

A FLAG-tagged construct and human influenza hemagglutinin (HA)-tagged constructs were co-transfected into rice protoplasts. After incubation for 14 h, transfected protoplasts were lysed in extraction buffer (100 mM Tris-HCl [pH 7.5], 150 mM NaCl, 5 mM EDTA, 5 mM EGTA, 2 mM DTT, 0.5% Triton X-100, and 1× Protease Inhibitor Cocktail [Roche]). After sonication, the samples were centrifuged at 12,000 × *g* for 10 min at 4°C; 100 μL of supernatant was used as input, and the remainder was incubated with anti-FLAG antibody (F3165; Sigma) for 3 h at 4°C. Protein G Agarose beads (Roche) were added to the mixture, and the sample was rotated for an additional 3 h. The bead-bound proteins were captured, washed three times, and then resolved by SDS–PAGE and blotted onto a polyvinylidene difluoride (PVDF) membrane (Millipore). The presence of Gnd5-HA protein was detected by anti-HA (H6908; Sigma) with ECL Reagent (GE Healthcare).

Bimolecular fluorescence complementation assay

To perform the BiFC assay, *Gnd5-nV or nV* were co-transformed with *Gnd5-cV*. *NLS-RFP* was also transformed as a nuclear marker. After incubation for 14 h, the yellow fluorescent protein (YFP) and red fluorescent protein (RFP) signals were detected using a fluorescence microscope. Similar results were observed in at least 50 cells from three independent experiments.

Yeast two-hybrid assay

To perform the yeast two-hybrid assay, the Matchmaker Gold Yeast Two-Hybrid System (Clontech, USA) was used. Full-length *Gnd5*, $Gnd5^{S185A+T420I}$, and $Gnd5^{S185A}$ were cloned into the *pGADT7* vector containing the GAL4 activation domain (AD) and the *pGBKT7* vector containing the GAL4 binding domain (BD). The AD and BD constructs were transformed into yeast strains Y187 and Y2HGold, respectively. All yeast transformants were grown on SD/-Trp/-Leu/-His/-Ade medium for selection or interaction tests in the presence of 40 g ml^{-1} X-α-Gal. Plates were incubated at 28°C for 3–4 d.

RNA isolation and RT-qPCR analysis

Total RNA was isolated from rice protoplasts using the TRIzol reagent (Invitrogen, catalog no. AM1912). RNA (2 μg) was used to prepare first-strand cDNA using TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix (TransGen Biotech, catalog no. AU311-02). RT-qPCR was performed using THUNDERBIRD SYBR qPCR Mix without Rox reagent (TOYOBO, catalog no. QPS-20(-)). Three biological replicates were performed for each gene. Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as the internal control. Primers for RT-qPCR are listed in Supplemental Table 15.

Estimation of the phenotypic variance explained by introgression and selection

Estimation of the phenotypic variance of eight agronomic traits explained by introgression and selection was performed by GREML (genome-based restricted maximum likelihood) analysis using Genome-wide Complex Trait Analysis (GCTA) software with default settings (Yang et al., 2011). In our analysis, the top three PCs were used as covariates, and variance of the genetic factor was determined by the genotypes of SNPs in the genomic introgression and selection regions.

Analysis of gene effects on grain number per panicle

We collected 51 cloned genes and the newly identified gene *Gnd5* related to GN and identified putative functional variations (mutations that altered the amino acids or were located upstream of the gene) in the 318 rice accessions with phenotype data. A subset of 33 candidates with functional variations was included to conduct gene effects analysis. To identify superior haplotypes, we referred to the QTN information provided by Wei et al. (2021). For some genes without QTN information, we analyzed the genetic effects of those genes that had at least two haplotypes and corresponding phenotype data. We used a t-test to identify significant phenotype differences between two major haplotypes under at least one environment when defining inferior and superior haplotypes (Li et al., 2020). The genetic effect of each gene was estimated separately on the basis of a linear regression model using the R package NOIA (Alvarez-Castro and Carlborg, 2007) in three pedigrees (Liaogeng5, Jigeng60, and Hejiang20) and locations (Liaoning, Jilin, and Heilongjiang).

Availability of data and materials

The raw sequence data reported in this paper have been deposited in the National Genomics Data Center (NGDC), part of the China National Center for Bioinformation (CNCB), under accession codes CRA004238 and CRA005267.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xplc.2022.100325.

FUNDING

This work was supported by the National Key Research and Development Program of China (2021YFD1200500 and 2016YFD0100101), National Key Research and Development Program of China (2016YFD0100801 and 2017YFA0503800), CAAS Science and Technology Innovation Program, Protective Program of Crop Germplasm of China (19200385-1), Third National Survey And Collection Action On Crop Germplasm Resource (19210859 and 19210860), National Crop Germplasm Resources Center (NCGRC-2021-02), and Project of Sanya Yazhou Bay Science and Technology City (SKJC-2020-02-001).

AUTHOR CONTRIBUTIONS

D.C., L.H., and H.H. carried out project design and material preparation. D.C. and X.M. were responsible for rice resequencing. H.Z. and D.C. contributed to data analyses. B.H., M.L., J.S., J.L., G.J., X.W., and G.C. contributed to field trials and phenotype collection. D.C., H.Z., Z.L., L.S., H.H., L.H., and X.W.D. contributed to manuscript writing. L.H. and H.H. conceived the idea, performed the data interpretation, and wrote the manuscript. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We thank the China National Crop Genebank for providing the rice seeds. We thank Biogle for producing transgenic rice. Most of the bioinformatic analysis was performed on the High Performance Computing Platform of the Center for Life Sciences (Peking University). No conflict of interest is declared.

Received: October 29, 2021
Revised: March 12, 2022
Accepted: April 9, 2022
Published: May 9, 2022

REFERENCES

Alvarez-Castro, J.M., and Carlborg, O. (2007). A unified model for functional and statistical epistasis and its application in quantitative trait Loci analysis. Genetics **176**:1151–1167.

Bai, X., Huang, Y., Hu, Y., Liu, H., Zhang, B., Smaczniak, C., Hu, G., Han, Z., and Xing, Y. (2017). Duplication of an upstream silencer of *FZP* increases grain yield in rice. Nat. Plants **3**:885–893.

Bai, S., Yu, H., Wang, B., and Li, J. (2018). Retrospective and perspective of rice breeding in China. J. Genet. Genomics **45**:603–612.

Bates, R., Battistin, M., Berry, S., et al. (2014). Implementation of ultrasonic sensing for high resolution measurement of binary gas mixture fractions. Sensors **14**:11260–11276.

Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. **81**:1084–1097.

Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., Rivas, S., Alaux, L., Kanzaki, H., Okuyama, Y., et al. (2013). The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. Plant Cell **25**:1463–1481.

Chen, X., Shang, J., Chen, D., Lei, C., Zou, Y., Zhai, W., Liu, G., Xu, J., Ling, Z., Cao, G., et al. (2006). A B-lectin receptor kinase gene conferring rice blast resistance. Plant J. **46**:794–804.

Chen, W., Xu, Z., Zhang, L., Zhang, W., and Ma, D. (2007). Theories and practices of breeding *japonica* rice for super high yield. Scientia Agricultura Sinica **40**:869–874.

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

Chen, W., Gao, Y., Xie, W., Gong, L., Lu, K., Wang, W., Li, Y., Liu, X., Zhang, H., Dong, H., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. Nat. Genet. **46**:714–721.

Chen, J., Gao, H., Zheng, X.M., Jin, M., Weng, J.F., Ma, J., Ren, Y., Zhou, K., Wang, Q., Wang, J., et al. (2015). An evolutionarily conserved gene, *FUWA*, plays a role in determining panicle architecture, grain shape and grain weight in rice. Plant J. **83**:427–438.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics **34**:i884–i890.

Chen, Z., Li, X., Lu, H., Gao, Q., Du, H., Peng, H., Qin, P., and Liang, C. (2020). Genomic atlases of introgression and differentiation reveal breeding footprints in Chinese cultivated rice. J. Genet. Genomics **47**:637–649.

Chu, Z., Fu, B., Yang, H., Xu, C., Li, Z., Sanchez, A., Park, Y.J., Bennetzen, J.L., Zhang, Q., and Wang, S. (2006). Targeting *xa13*, a recessive gene for bacterial blight resistance in rice. Theor. Appl. Genet. **112**:455–461.

Cingolani, P., Platts, A., Wang le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (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 (Austin) **6**:80–92.

Cubry, P., Tranchant-Dubreuil, C., Thuillet, A.C., Monat, C., Ndjiondjop, M.N., Labadie, K., Cruaud, C., Engelen, S., Scarcelli, N., Rhoné, B., et al. (2018). The rise and fall of african rice cultivation revealed by analysis of 246 new genomes. Curr. Biol. **14**:2274–2282.

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.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**:15–21.

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.

Fitak, R.R. (2021). OptM: estimating the optimal number of migration edges on population trees using Treemix. Biol. Methods Protoc. **6**:bpab017.

Fitzgerald, M.A., McCouch, S.R., and Hall, R.D. (2009). Not just a grain of rice: the quest for quality. Trends in Plant Science **14**:133–139.

Fu, F., and Xue, H. (2010). Coexpression analysis identifies Rice Starch Regulator1, a rice AP2/EREBP family transcription factor, as a novel rice starch biosynthesis regulator. Plant Physiol. **154**:927–938.

Gao, Z., Zeng, D., Cui, X., Zhou, Y., Yan, M., Huang, D., Li, J., and Qian, Q. (2003). Map-based cloning of the *ALK* gene, which controls the gelatinization temperature of rice. Sci. China C Life Sci. **46**:661–668.

Gao, H., Jin, M., Zheng, X.M., Chen, J., Yuan, D., Xin, Y., Wang, M., Huang, D., Zhang, Z., Zhou, K., et al. (2014). *Days to heading 7*, a major quantitative locus determining photoperiod sensitivity and regional adaptation in rice. Proc. Natl. Acad. Sci. U S A **111**:16337–16342.

Gao, Z., Wang, Y., Chen, G., Zhang, A., Yang, S., Shang, L., Wang, D., Ruan, B., Liu, C., Jiang, H., et al. (2019). The *indica* nitrate reductase gene *OsNR2* allele enhances rice yield potential and nitrogen use efficiency. Nat. Commun. **10**:5207.

Garg, R., Jhanwar, S., Tyagi, A.K., and Jain, M. (2010). Genome-wide survey and expression analysis suggest diverse roles of glutaredoxin gene family members during development and response to various stimuli in rice. DNA Res. **17**:353–367.

Genomes project. (2014). The 3,000 rice genomes project. Gigascience **3**:7.

Gu, H. (2010). Discussion on the aspects of high-yielding breeding in rice. ACTA Agronomica Sinica **36**:1431–1439.

Guo, S., Xu, Y., Liu, H., Mao, Z., Zhang, C., Ma, Y., Zhang, Q., Meng, Z., and Chong, K. (2013). The interaction between *OsMADS57* and *OsTB1* modulates rice tillering via *DWARF14*. Nat. Commun. **4**:1566.

Hakoshima, T. (2018). Structural basis of the specific interactions of GRAS family proteins. FEBS Lett. **4**:489–501.

1. **Han, L., Wei, X., Cao, G.L., Yu, H.Y., and Zhang, Y.Y.** (2006). Descriptors and Data Standard for Rice (Oryza sativa L.), 4 (Beijing: China Agriculture Press).

Hao, C., Jiao, C., Hou, J., Li, T., Liu, H., Wang, Y., Zheng, J., Liu, H., Bi, Z., Xu, F., et al. (2020). Resequencing of 145 landmark cultivars reveals asymmetric sub-genome selection and strong founder genotype effects on wheat breeding in China. Mol. Plant **13**:1733–1751.

Hu, B., Wang, W., Ou, S., Tang, J., Li, H., Che, R., Zhang, Z., Chai, X., Wang, H., Wang, Y., et al. (2015a). Variation in *NRT1.1B* contributes to nitrate-use divergence between rice subspecies. Nat. Genet. **47**:834–838.

Hu, J., Wang, Y., Fang, Y., Zeng, L., Xu, J., Yu, H., Shi, Z., Pan, J., Zhang, D., Kang, S., et al. (2015b). A rare allele of *GS2* enhances grain size and grain yield in rice. Mol. Plant **8**:1455–1465.

Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. Nat. Genet. **42**:961–967.

Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., Li, W., Guo, Y., Deng, L., Zhu, C., et al. (2011). Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat. Genet. **44**:32–39.

Huang, X., Kurata, N., Wei, X., Wang, Z.X., Wang, A., Zhao, Q., Zhao, Y., Liu, K., Lu, H., Li, W., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. Nature **490**:497–501.

Hufford, M.B., Xu, X., van Heerwaarden, J., Pyhajarvi, T., Chia, J.M., Cartwright, R.A., Elshire, R.J., Glaubitz, J.C., Guill, K.E., Kaeppler, S.M., et al. (2012). Comparative population genomics of maize domestication and improvement. Nat. Genet. **44**:808–811.

Hyten, D.L., Song, Q., Zhu, Y., Choi, I.Y., Nelson, R.L., Costa, J.M., Specht, J.E., Shoemaker, R.C., and Cregan, P.B. (2006). Impacts of genetic bottlenecks on soybean genome diversity. Proc. Natl. Acad. Sci. U S A **103**:16666–16671.

Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., Dong, G., Zeng, D., Lu, Z., Zhu, X., et al. (2010). Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. Nat. Genet. **42**:541–544.

Jin, J., Hua, L., Zhu, Z., Tan, L., Zhao, X., Zhang, W., Liu, F., Fu, Y., Cai, H., Sun, X., et al. (2016). *GAD1* encodes a secreted peptide that regulates grain number, grain length, and awn development in rice domestication. Plant Cell **28**:2453–2463.

Kamolsukyunyong, W., Sukhaket, W., Ruanjaichon, V., Toojinda, T., and Vanavichit, A. (2013). Single-feature polymorphism mapping of isogenic rice lines identifies the influence of terpene synthase on brown planthopper feeding preferences. Rice (N Y) **1**:18.

Kang, H.M., Sul, J.H., Service, S.K., Zaitlen, N.A., Kong, S.Y., Freimer, N.B., Sabatti, C., and Eskin, E. (2010). Variance component model to account for sample structure in genome-wide association studies. Nat. Genet. **42**:348–354.

Kawahara, Y., de la Bastide, M., Hamilton, J.P., Kanamori, H., McCombie, W.R., Ouyang, S., Schwartz, D.C., Tanaka, T., Wu, J., Zhou, S., et al. (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. Rice (N Y) **6**:4.

Komatsu, M., Maekawa, M., Shimamoto, K., and Kyozuka, J. (2001). The *LAX1* and *FRIZZY PANICLE 2* genes determine the inflorescence architecture of rice by controlling rachis-branch and spikelet development. Dev. Biol. **231**:364–373.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. **33**:1870–1874.

Leigh, J.W., Bryant, D., Nakagawa, S.J.M.i.E., and Evolution. (2015). Popart : full-feature software for haplotype network construction. Methods Ecol. Evol. **6**:1110–1116.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics **21**:2987–2993.

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., Durbin, R., and Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. Bioinformatics **25**:2078–2079.

Li, F., Liu, W., Tang, J., Chen, J., Tong, H., Hu, B., Li, C., Fang, J., Chen, M., and Chu, C. (2010). Rice *DENSE AND ERECT PANICLE 2* is essential for determining panicle outgrowth and elongation. Cell Res **20**:838–849.

Li, J., Sun, Y., Liu, H., Wang, Y., Jia, Y., and Xu, M. (2015). Natural variation of rice blast resistance gene *Pi-d2*. Genet. Mol. Res. **14**:1235–1249.

Li, X., Chen, Z., Zhang, G., Lu, H., Qin, P., Qi, M., Yu, Y., Jiao, B., Zhao, X., Gao, Q., et al. (2020). Analysis of genetic architecture and favorable allele usage of agronomic traits in a large collection of Chinese rice accessions. Sci. China Life Sci. **63**:1688–1702.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics **30**:923–930.

Liu, Q., Yuan, M., Zhou, Y., Li, X., Xiao, J., and Wang, S. (2011). A paralog of the MtN3/saliva family recessively confers race-specific resistance to *Xanthomonas oryzae* in rice. Plant Cell Environ. **34**:1958–1969.

Liu, Y., Liu, B., Zhu, X., Yang, J., Bordeos, A., Wang, G., Leach, J.E., and Leung, H. (2013). Fine-mapping and molecular marker development for *Pi56(t)*, a NBS-LRR gene conferring broad-spectrum resistance to *Magnaporthe oryzae* in rice. Theor. Appl. Genet. **126**:985–998.

Liu, J., Chen, J., Zheng, X., Wu, F., Lin, Q., Heng, Y., Tian, P., Cheng, Z., Yu, X., Zhou, K., et al. (2017). *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nat. Plants **3**:17043.

Liu, M., Zhang, S., Hu, J., Sun, W., Padilla, J., He, Y., Li, Y., Yin, Z., Liu, X., Wang, W., et al. (2019). Phosphorylation-guarded light-harvesting complex II contributes to broad-spectrum blast resistance in rice. Proc. Natl. Acad. Sci. U S A **116**:17572–17577.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**:550.

Ma, Y., Dai, X., Xu, Y., Luo, W., Zheng, X., Zeng, D., Pan, Y., Lin, X., Liu, H., Zhang, D., et al. (2015). *COLD1* confers chilling tolerance in rice. Cell **160**:1209–1221.

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.

McNally, K.L., Childs, K.L., Bohnert, R., Davidson, R.M., Zhao, K., Ulat, V.J., Zeller, G., Clark, R.M., Hoen, D.R., Bureau, T.E., et al. (2009). Genomewide SNP variation reveals relationships among landraces and modern varieties of rice. Proc. Natl. Acad. Sci. U S A **106**:12273–12278.

Mizuta, Y., Harushima, Y., and Kurata, N. (2010). Rice pollen hybrid incompatibility caused by reciprocal gene loss of duplicated genes. Proc. Natl. Acad. Sci. U S A **107**:20417–20422.

Murray, M.G., and Thompson, W.F. (1980). Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res. **19**:4321–4325.

Nemoto, Y., Hori, K., and Izawa, T. (2018). Fine-tuning of the setting of critical day length by two casein kinases in rice photoperiodic flowering. J. Exp. Bot. **69**:553–565.

Niu, X., Chen, S., Li, J., Liu, Y., Ji, W., and Li, H. (2019). Genome-wide identification of GRAS genes in *Brachypodium distachyon* and functional characterization of *BdSLR1* and *BdSLRL1*. BMC Genomics **20**:635.

Okuyama, Y., Kanzaki, H., Abe, A., Yoshida, K., Tamiru, M., Saitoh, H., Fujibe, T., Matsumura, H., Shenton, M., Galam, D.C., et al. (2011). A multifaceted genomics approach allows the isolation of the rice *Pia*-blast resistance gene consisting of two adjacent NBS-LRR protein genes. Plant J. **3**:467–479.

Pedersen, B.S., and Quinlan, A.R. (2018). Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics **5**:867–868.

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

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. **81**:559–575.

Qiao, Y., Piao, R., Shi, J., Lee, S.I., Jiang, W., Kim, B.K., Lee, J., Han, L., Ma, W., and Koh, H.J. (2011). Fine mapping and candidate gene analysis of *dense and erect panicle 3, DEP3*, which confers high grain yield in rice (*Oryza sativa* L.). Theor. Appl. Genet. **122**:1439–1449.

Ren, J., Gao, F., Wu, X., Lu, X., Zeng, L., Lv, J., Su, X., Luo, H., and Ren, G. (2016). *Bph32*, a novel gene encoding an unknown SCR domain-containing protein, confers resistance against the brown planthopper in rice. Sci. Rep. **6**:37645.

Santos, J.D., Chebotarov, D., McNally, K.L., Bartholome, J., Droc, G., Billot, C., and Glaszmann, J.C. (2019). Fine scale genomic signals of admixture and alien introgression among asian rice landraces. Genome Biol. Evol. **11**:1358–1373.

Song, W.Y., Wang, G.L., Chen, L.L., Kim, H.S., Pi, L.Y., Holsten, T., Gardner, J., Wang, B., Zhai, W.X., Zhu, L.H., et al. (1995). A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa21*. Science **270**:1804–1806.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **30**:1312–1313.

Sun, J., Liu, D., Wang, J.Y., Ma, D.R., Tang, L., Gao, H., Xu, Z.J., and Chen, W.F. (2012). The contribution of intersubspecific hybridization to the breeding of super-high-yielding *japonica* rice in northeast China. Theor. Appl. Genet. **125**:1149–1157.

Takahashi, A., Hayashi, N., Miyao, A., and Hirochika, H. (2010). Unique features of the rice blast resistance *Pish* locus revealed by large scale retrotransposon-tagging. BMC Plant Biol. **10**:175.

Takeda, T., Suwa, Y., Suzuki, M., Kitano, H., Ueguchi-Tanaka, M., Ashikari, M., Matsuoka, M., and Ueguchi, C. (2003). The *OsTB1* gene negatively regulates lateral branching in rice. Plant J. **33**:513–520.

Tamura, Y., Hattori, M., Yoshioka, H., Yoshioka, M., Takahashi, A., Wu, J., Sentoku, N., and Yasui, H. (2014). Map-based cloning and characterization of a brown planthopper resistance gene *BPH26* from *Oryza sativa* L. ssp. *indica* cultivar ADR52. Sci. Rep. **4**:5872.

Tan, L., Li, X., Liu, F., Sun, X., Li, C., Zhu, Z., Fu, Y., Cai, H., Wang, X., Xie, D., et al. (2008). Control of a key transition from prostrate to erect growth in rice domestication. Nat. Genet. **40**:1360–1364.

Tang, L., and Chen, W. (2021). Development trend and prospect of Geng rice in northeast China. China Rice **05**:1–4.

Tian, C., Wan, P., Sun, S., Li, J., and Chen, M. (2004). Genome-wide analysis of the GRAS gene family in rice and *Arabidopsis*. Plant Molecular Biology **54**:519–532.

Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., and Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. Nucleic Acids Res. **45**:W122–W129.

Tong, H., Liu, L., Jin, Y., Du, L., Yin, Y., Qian, Q., Zhu, L., and Chu, C. (2012). Dwarf and low-tillering acts as a direct downstream target of a GSK3/SHAGGY-like kinase to mediate brassinosteroid responses in rice. Plant Cell **6**:2562–2577.

Vasudevan, K., Vera Cruz, C.M., Gruissem, W., and Bhullar, N.K. (2016). Geographicly distinct and domain-specific sequence variations in the alleles of rice blast resistance gene *Pib*. Front. Plant Sci. **7**:915.

Wang, X., Cheng, K., Huang, N., Luo, J., Lu, Y., and Liu, G. (1987). Studies on two important rice subspecies types during origin and evolution of Asian cultivated rice. Acta Genetica Sinica **4**:262–270.

Wang, L., Xu, Y., Zhang, C., Ma, Q., Joo, S.H., Kim, S.K., Xu, Z., and Chong, K. (2008). OsLIC, a novel CCCH-type zinc finger protein with transcription activation, mediates rice architecture via brassinosteroids signaling. PLoS One **3**:e3521.

Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R., Zhang, F., et al. (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature **557**:43–49.

Wei, X., Qiu, J., Yong, K., Fan, J., Zhang, Q., Hua, H., Liu, J., Wang, Q., Olsen, K.M., Han, B., et al. (2021). A quantitative genomics map of rice provides genetic insights and guides breeding. Nat. Genet. **53**:243–253.

Wu, W., Zheng, X.M., Lu, G., Zhong, Z., Gao, H., Chen, L., Wu, C., Wang, H.J., Wang, Q., Zhou, K., et al. (2013). Association of functional nucleotide polymorphisms at *DTH2* with the northward expansion of rice cultivation in Asia. Proc. Natl. Acad. Sci. U S A **110**:2775–2780.

Wu, Y., Wang, Y., Mi, X.F., Shan, J.X., Li, X.M., Xu, J.L., and Lin, H.X. (2016). The QTL *GNP1* encodes GA20ox1, which increases grain number and yield by increasing cytokinin activity in rice panicle meristems. PLoS Genet. **12**:e1006386.

Xiao, N., Pan, C., Li, Y., Wu, Y., Cai, Y., Lu, Y., Wang, R., Yu, L., Shi, W., Kang, H., et al. (2021). Genomic insight into balancing high yield, good quality, and blast resistance of *japonica* rice. Genome Biol. **22**:283.

Xie, W., Wang, G., Yuan, M., Yao, W., Lyu, K., Zhao, H., Yang, M., Li, P., Zhang, X., Yuan, J., et al. (2015). Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. Proc. Natl. Acad. Sci. U S A **112**:E5411–E5419.

Xing, Y., and Zhang, Q. (2010). Genetic and molecular bases of rice yield. Annu. Rev. Plant Biol. **61**:421–442.

Xu, X., Liu, X., Ge, S., Jensen, J.D., Hu, F., Li, X., Dong, Y., Gutenkunst, R.N., Fang, L., Huang, L., et al. (2011). Resequencing 50 accessions of cultivated and wild rice yields markers for identifying agronomicly important genes. Nat. Biotechnol. **30**:105–111.

Xu, R., Duan, P., Yu, H., Zhou, Z., Zhang, B., Wang, R., Li, J., Zhang, G., Zhuang, S., Lyu, J., et al. (2018). Control of grain size and weight by the OsMKKK10-OsMKK4-OsMAPK6 signaling pathway in rice. Mol. Plant **11**:860–873.

Xue, W., Xing, Y., Weng, X., Zhao, Y., Tang, W., Wang, L., Zhou, H., Yu, S., Xu, C., Li, X., et al. (2008). Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. Nat. Genet. **40**:761–767.

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.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.C., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H., Hirano, K., et al. (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat. Genet. **48**:927–934.

Yu, J., Xiong, H., Zhu, X., Zhang, H., Li, H., Miao, J., Wang, W., Tang, Z., Zhang, Z., Yao, G., et al. (2017). *OsLG3* contributing to rice grain length and yield was mined by Ho-LAMap. BMC Biol. **15**:28.

Zhang, Q., Li, J., Zhang, W., Yan, S., Wang, R., Zhao, J., Li, Y., Qi, Z., Sun, Z., and Zhu, Z. (2012). The putative auxin efflux carrier *OsPIN3t* is involved in the drought stress response and drought tolerance. Plant J. **72**:805–816.

Zhang, H., Zhang, J., Chang, Y., Li, M., Gao, H., Dai, Q., Huo, Z., Xu, K., and Wei, H. (2013). The productive advantages and formation mechanisms of "*indica* rice to *japonica* rice. Scientia Agricultura Sinica **46**:686–704.

Zhang, G.H., Li, S.Y., Wang, L., Ye, W.J., Zeng, D.L., Rao, Y.C., Peng, Y.L., Hu, J., Yang, Y.L., Xu, J., et al. (2014). *LSCHL4* from *Japonica* Cultivar, which is allelic to *NAL1*, increases yield of *indica* super rice 93-11. Mol. Plant **7**:1350–1364.

Zhang, Y., Tan, L., Zhu, Z., Yuan, L., Xie, D., and Sun, C. (2015). *TOND1* confers tolerance to nitrogen deficiency in rice. Plant J. **81**:367–376.

Zhang, F., Wang, C., Li, M., Cui, Y., Shi, Y., Wu, Z., Hu, Z., Wang, W., Xu, J., and Li, Z. (2021). The landscape of gene-CDS-haplotype diversity in rice: properties, population organization, footprints of domestication and breeding, and implications for genetic improvement. Mol. Plant **14**:787–804.

Zhao, K., Wright, M., Kimball, J., Eizenga, G., McClung, A., Kovach, M., Tyagi, W., Ali, M.L., Tung, C.W., Reynolds, A., et al. (2010). Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. PLoS One **5**:e10780.

Zhou, H., Alexander, D., and Lange, K. (2011). A quasi-Newton acceleration for high-dimensional optimization algorithms. Stat. Comput. **21**:261–273.

Zhou, J., Cui, D., Zhao, S., Sun, J., Lo, M., Ma, X., Mao, X., Li, X., Zhao, J., Piao, D., et al. (2019). Phenotypic evaluation of main agronomic traits in rice varieties that were released over the past few decades. J. Plant Genet. Resour. **20**:1566–1578.