SCIENCE CHINA
Life Sciences

•RESEARCH PAPER•

November 2020 Vol.63 No.11: 1688–1702
https://doi.org/10.1007/s11427-019-1682-6

Analysis of genetic architecture and favorable allele usage of agronomic traits in a large collection of Chinese rice accessions

Xiuxiu Li[1,2†], Zhuo Chen[1,2†], Guomin Zhang[3†], Hongwei Lu[1,2†], Peng Qin[4], Ming Qi[1], Ying Yu[1], Bingke Jiao[1], Xianfeng Zhao[1], Qiang Gao[1], Hao Wang[4], Yunyu Wu[5,6], Juntao Ma[3], Liyan Zhang[3], Yongli Wang[3], Lingwei Deng[3], Shanguo Yao[1], Zhukuang Cheng[1], Diqiu Yu[7], LihuangZhu[1], Yongbiao Xue[1], Chengcai Chu[1], Aihong Li[5,6*], Shigui Li[4*] & Chengzhi Liang[1,2*]

[1]*State Key Laboratory of Plant Genomics, Institute of Genetics and Developmental Biology, Innovation Academy for Seed Design, Chinese Academy of Sciences, Beijing 100101, China;*
[2]*University of Chinese Academy of Sciences, Beijing 100049, China;*
[3]*Biotechnology Research Institute, Heilongjiang Academy of Agricultural Sciences, Harbin 150086, China;*
[4]*Rice Research Institute, State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Sichuan Agricultural University, Chengdu 611130, China;*
[5]*Lixiahe Agricultural Research Institute of Jiangsu Province, Yangzhou 225009, China;*
[6]*Jiangsu Collaborative Innovation Center for Modern Crop Production, Nanjing 210095, China;*
[7]*State Key Laboratory for Conservation and Utilization of Bio-Resources in Yunnan, Yunnan University, Kunming 650091, China*

Received January 15, 2020; accepted March 16, 2020; published online April 15, 2020

Genotyping and phenotyping large natural populations provide opportunities for population genomic analysis and genome-wide association studies (GWAS). Several rice populations have been re-sequenced in the past decade; however, many major Chinese rice cultivars were not included in these studies. Here, we report large-scale genomic and phenotypic datasets for a collection mainly comprised of 1,275 rice accessions of widely planted cultivars and parental hybrid rice lines from China. The population was divided into three *indica*/*Xian* and three *japonica*/*Geng* phylogenetic subgroups that correlate strongly with their geographic or breeding origins. We acquired a total of 146 phenotypic datasets for 29 agronomic traits under multi-environments for different subpopulations. With GWAS, we identified a total of 143 significant association loci, including three newly identified candidate genes or alleles that control heading date or amylose content. Our genotypic analysis of agronomically important genes in the population revealed that many favorable alleles are underused in elite accessions, suggesting they may be used to provide improvements in future breeding efforts. Our study provides useful resources for rice genetics research and breeding.

rice, Chinese cultivars, whole-genome resequencing, multi-environmental phenotyping, genome-wide association studies, favorable alleles

Citation: 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. https://doi.org/10.1007/s11427-019-1682-6

INTRODUCTION

Cultivated Asian rice (*Oryza sativa* L.) is one of the most important staple food crops (Xing and Zhang, 2010). Utilization of semi-dwarfness and exploitation of hybrid vigor have greatly increased rice productivity in China over the past several decades. Currently, China is a major center for rice production, research, and breeding efforts. Each year, many cultivars are released that improve production, grain

†Contributed equally to this work
*Corresponding authors (Chengzhi Liang, email: cliang@genetics.ac.cn; Shigui Li, email: lishigui_sc@263.net; Aihong Li, email: yzlah@126.com)

quality, and disease resistance.

Recent advancement of next-generation sequencing technologies has enabled the sequencing of numerous rice accessions at relatively low cost, which provides opportunities for large-scale population genetics and genomic analyses. In the last decade, genotyping datasets for many rice accessions have been published, from which a large number of loci associated with important agronomic traits have been identified (Crowell et al., 2016; Dong et al., 2018; Huang et al., 2010; Huang et al., 2015; Huang et al., 2011; Li et al., 2019; Wang et al., 2015b; Zhao et al., 2018). It has been found that several of these loci were selected during rice domestication and breeding (Huang et al., 2012; Xie et al., 2015). Recently, the 3,000 Rice Genomes Project has provided a large-scale genotyping resource for rice genomic research (Wang et al., 2018b). Meanwhile, substantial phenotyping data obtained via traditional and high-throughput phenotyping technologies have also been reported (Crowell et al., 2016; Guo et al., 2018; Huang et al., 2010; Huang et al., 2011; Yang et al., 2014). In crop plants, including rice, genome-wide association studies (GWAS) have resulted in remarkable discoveries about the genetic basis of complex agronomic traits, resistant characteristics, and metabolites (Chen et al., 2014; Crowell et al., 2016; Dong et al., 2018; Duan et al., 2017; Guo et al., 2018; Huang et al., 2010; Huang et al., 2011; Wang et al., 2015b; Yang et al., 2014). Thus, GWAS has been a powerful tool for exploring the genetic basis of important rice traits. In particular, GWAS with multi-trait and multi-environmental phenotyping can facilitate our understanding of genotype-phenotype relationships and genotype-environment interactions. Nevertheless, well-studied rice accessions mostly represent traditional varieties and landraces that were collected from many countries but do not fully represent current Chinese rice cultivars which resulted from intensive modern breeding efforts and carry many agronomically favorable alleles.

In this study, we report large-scale genomic and phenomic datasets for a collection of 1,275 rice accessions that mainly includes Chinese rice cultivars and parental lines. We sequenced these rice accessions as Illumina platform short reads (Illumina Inc, USA). We collected 146 phenotypic datasets on multi-traits under multi-environments and detected nine GWAS signals linked to candidate genes with known and unknown functions. We analyzed the usage pattern of favorable alleles of agronomically important genes. We found that while many favorable alleles are widely used in elite lines, the favorable alleles of some genes are still underused and could be further exploited in future breeding programs. Our genomic and phenomic datasets will facilitate rice genetics research and elite cultivar breeding for sustainable agriculture.

RESULTS

SNP identification

Our diverse collection of 1,275 rice accessions is comprised of Chinese rice cultivars, parental lines of hybrid rice cultivars, and a small number of elite rice accessions that have been imported into China (Table S1 in Supporting Information). The accessions were sequenced to an average of 7.2× depth with an Illumina 125–150 bp paired-end reads protocol. Short reads were mapped to the Nipponbare reference genome (MSU version 7.0), averaging 89.3% genome coverage. After SNP calling, we identified a total of 2.5 million high-quality SNPs that had a minor allele frequency greater than 0.01, at an average of 6.8 SNPs per kb (Table 1). We found 397,207 (15.6%), 583,820 (23.0%), 125,427 (4.9%), and 1,662,909 (65.5%) SNPs located within exons, introns, UTRs, and intergenic regions, respectively (Table 1;

Table 1 Summary of genomic variants in different rice populations

	All	XI	GJ
Number of accessions	1,275	422	834
Total SNPs	2,538,380	2,412,486	2,257,656
SNPs with population-specific alleles[a)]	–	371,454	374,706
SNPs in exon	397,207	376,617	356,892
SNPs in intron	583,820	583,820	518,639
SNPs in UTR	125,427	119,497	109,626
SNPs in intergenic region	1,662,909	1,581,049	1,476,772
Non-synonymous SNPs	223,126	210,710	200,278
Splicing SNPs	2,506	2,369	2,250
Start-gain SNPs	6,504	6,195	5,726
Stop-gain SNPs	8,769	8,181	7,793
Stop-loss SNPs	1,074	1,037	1,012

a) SNPs with population-specific alleles are those with an allele frequency >0.95 in one subpopulation and <0.05 in another.

Figure S1B and C in Supporting Information), including 223,126 non-synonymous, 2,506 splice site, 6,504 start-gain, 8,769 stop-gain, and 1,074 stop-loss SNPs in coding regions (Table 1). After filtering (for missing data less than 0.2 and minor allele frequency greater than 0.05), we selected 2,081,216 SNPs for use in subsequent analyses.

Population genomic analysis

The 1,275 rice accessions were mainly divided into two distinct populations, Xian/*indica* (XI) and Geng/*japonica* (GJ) (Wang et al., 2018b) (Figure 1A–C), which included 413 typical XI and 818 typical GJ accessions (Table S1 in Supporting Information). We found that the allele frequencies of over 40% of the SNPs were very different between XI and GJ (with a difference greater than 0.7), including 746,160 population-specific alleles with a frequency difference greater than 0.9 (Table 1).

By K-means clustering of the first three principal components and phylogenetic tree analysis, the XI and GJ populations were further divided into three groups each (Figure 1A–C). The population structure correlated well with accession pedigree and geographic distribution in China (Table S3 in Supporting Information). GJ groups 1 and 2 were mainly early maturing rice accessions distributed in the Northeast of China, with group 1 including many varieties derived from several elite parental lines and group 2 mainly comprised of landraces and imported varieties. GJ group 3 mainly included medium and late maturing rice varieties derived from Wuyujing 3, Zhendao 88, Nanjing 11, and Xiushui series that are located in Northern, Central, and Southern China, particularly along the middle and lower stretches of the Yangtze River. XI group 1 included many hybrid rice maintainers derived from Zhanshan97, Xieqingzao, and FeigaiB. XI group 2 mainly contained BG90-2, Teqing, or Molizhan offspring that are located along the middle and lower Yangtze River. XI group 3 included many hybrid rice restorers and varieties derived from Gui630 and Minghui63 mainly located around the upper and middle Yangtze River. We found sterile hybrid rice lines scattered in

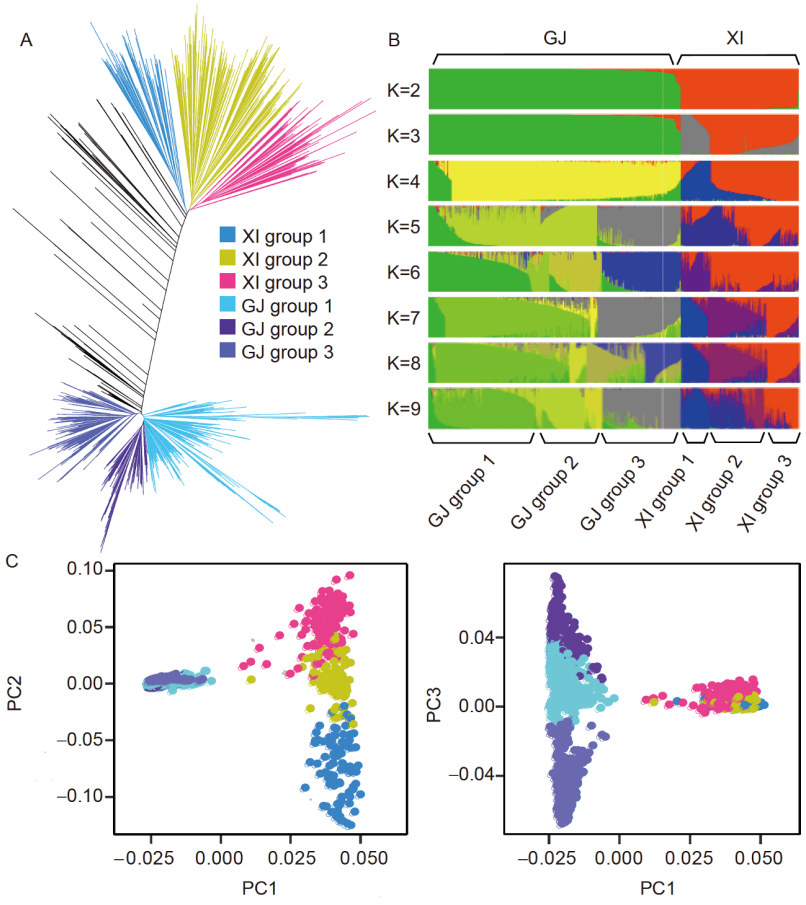

Figure 1 Population analysis of 1,275 rice accessions. A, Neighbor-joining tree of 1,275 rice accessions. The whole tree is divided into three XI subgroups and three GJ subgroups, which are color-coded with the same colors in A, C, D, and E. B, Population structure analysis of 1,275 rice accessions. Each color represents one ancestral subpopulation. Each accession is represented by a vertical bar, and the length of each colored segment in each vertical bar represents the proportion of genomic components contributed by an ancestral subpopulation. C, PCA plots of the first three components of the 1,275 rice accessions. PC1, PC2, and PC3 are abbreviations for the first three principle components. D, Geographic origin of the XI accessions. E, Geographic origin of the GJ accessions.

the XI subgroups.

We analyzed linkage disequilibrium (LD, indicated by r^2) value distribution patterns for the XI and GJ populations (Figure S2 in Supporting Information). We found LD dropped to half of its maximum value at 113 kb (as LD decay, r^2=0.28) in the whole population, but there was a large difference between the two subspecies (Figure S2A in Supporting Information). The XI population had a shorter LD decay distance (~124 kb, r^2=0.33) than the GJ population (~318 kb, r^2=0.40), consistent with the larger genetic diversity of the former over the latter. The LD decay of the XI population is consistent with a previous estimate (Huang et al., 2010). The LD decay in the GJ groups exhibited variations (Figure S2B in Supporting Information). The LD decay in GJ groups 1 and 3 increased to 624 kb and 638 kb, respectively, confirming the close relationship of GJ accessions resulting from intensive modern breeding selection that has deceased genetic diversity. We found that the LD varied in different genomic regions within both XI and GJ populations and between the two populations (Figure S2C in Supporting Information). These results suggest that reduced genetic diversity was produced in different genomic regions for the two populations during breeding selection.

Multi-environmental phenotyping in agro-ecologically diverse locations

We collected phenotypic data for the rice accessions described above from samples that were grown in several different regions of China (Tables S2 and S3 in Supporting Information). Due to insufficient seed or the unavailability of suitable planting areas, every accession was not planted in all regions or all seasons. In total, we obtained data for 1,159 rice accessions that were phenotyped for at least one season in at least one location.

A major subpopulation suited for growth (able to flower) in Northeast China included 450 GJ accessions that were mainly comprised of early maturing rice accessions from GJ groups 1 and 2 (hereafter called the NE-GJ subpopulation or panel). The rest of the population consisted of 709 rice varieties and breeding lines widely cultivated in China, including 299 typical XI and 369 typical GJ accessions (hereafter called the CN-Mix subpopulation or panel). Rice samples were planted at a total of 10 agro-ecologically diverse locations (Beijing, Jilin, Yangzhou, Wenjiang, Lingshui, and five cities in four temperate zones of Heilongjiang province) for measuring 29 agronomic traits in six categories: heading date; plant architecture (tiller number, panicle length, and plant height); yield (grain number, grain yield per plant, and others); grain size (grain length, grain width, and grain weight); grain appearance; and grain cooking characteristics (Figure 2; Tables S2–S5 in Supporting Information). In total, 146 phenotypes were measured, and 13 and 7 traits were measured at more than two locations for the NE-GJ and CN-Mix panels, respectively (Figure 2; Tables S4 and S5 in Supporting Information).

We found considerable phenotypic variations for most agronomic traits at different locations (Figure 2; Tables S4 and S5 in Supporting Information). The average days-to-heading (DTH) was lower in the South than in the North, which clearly results from the distinct photoperiods of these regions. For example, CN-Mix rice accessions planted at Lingshui (18.5°N, 110.0°E) have a shorter DTH on average than those at Beijing (39.9°N, 116.4°E) (89 at Lingshui vs. 113 at Beijing), while NE-GJ rice accessions planted at Heihe (50.2°N, 127.5°E) have the longest DTH (110 at Heihe vs. 65–102 at other locations). Plant height (PH) was positively correlated with DTH, with Pearson's correlation (r^2) being 0.24–0.55 and 0.39–0.67 for NE-GJ and CN-Mix, respectively. Grain yield per plant (GYP) and tiller number per plant (TNP) were also affected by environmental factors, showing large variations between different regions (r^2 between different locations were on average 0.14 and 0.18 for GYP and 0.33 and 0.32 for TNP for NE-GJ and CN-Mix, respectively). In contrast, grain length (GL) and grain width (GW) were found to be much less variable over all locations (r^2 between locations were on average 0.72 and 0.94 for GL, 0.54 and 0.90 for GW for NE-GJ and CN-Mix, respectively; coefficient of variation among all locations was 2.9% and 3.0% for GL, 3.7% and 2.4% for GW for NE-GJ and CN-Mix, respectively), suggesting that these traits are more genetically determined.

We examined pairwise correlations between different locations for all traits measured in multi-environments (Table S6 in Supporting Information) and found a high correlation (r^2>0.5) for most traits, including grain amylose content, grain protein content, grain length, grain width, plant height, and heading date. Notably, plant height varies significantly between different environments but has high correlation values (r^2 0.41–0.81 for NE-GJ and 0.56–0.88 for CN-Mix), suggesting linear genetic and environmental (GxE) interactions and a large additive genetic source for this trait (Bai et al., 2010; Zhang et al., 2010). We found a high correlation (r^2 0.48–0.89) for the heading date of the NE-GJ panel for each pair of locations. However, the heading date of the CN-Mix panel at Lingshui was weakly correlated with the other locations (r^2 0.03–0.48). Since HD is controlled by large effects from several genes that have differential sensitivity to photoperiod (Doi et al., 2004; Gao et al., 2014; Kovi et al., 2013; Song et al., 2012; Yano et al., 2000), this phenomenon may be the result of differential photoperiod sensitivities of the two populations. Relatively low correlation values between locations were observed for yield components GYP (r^2 0.03–0.28 and 0.05–0.32 for NE-GJ and CN-Mix, respectively), indicating the large effects the environment has on these traits.

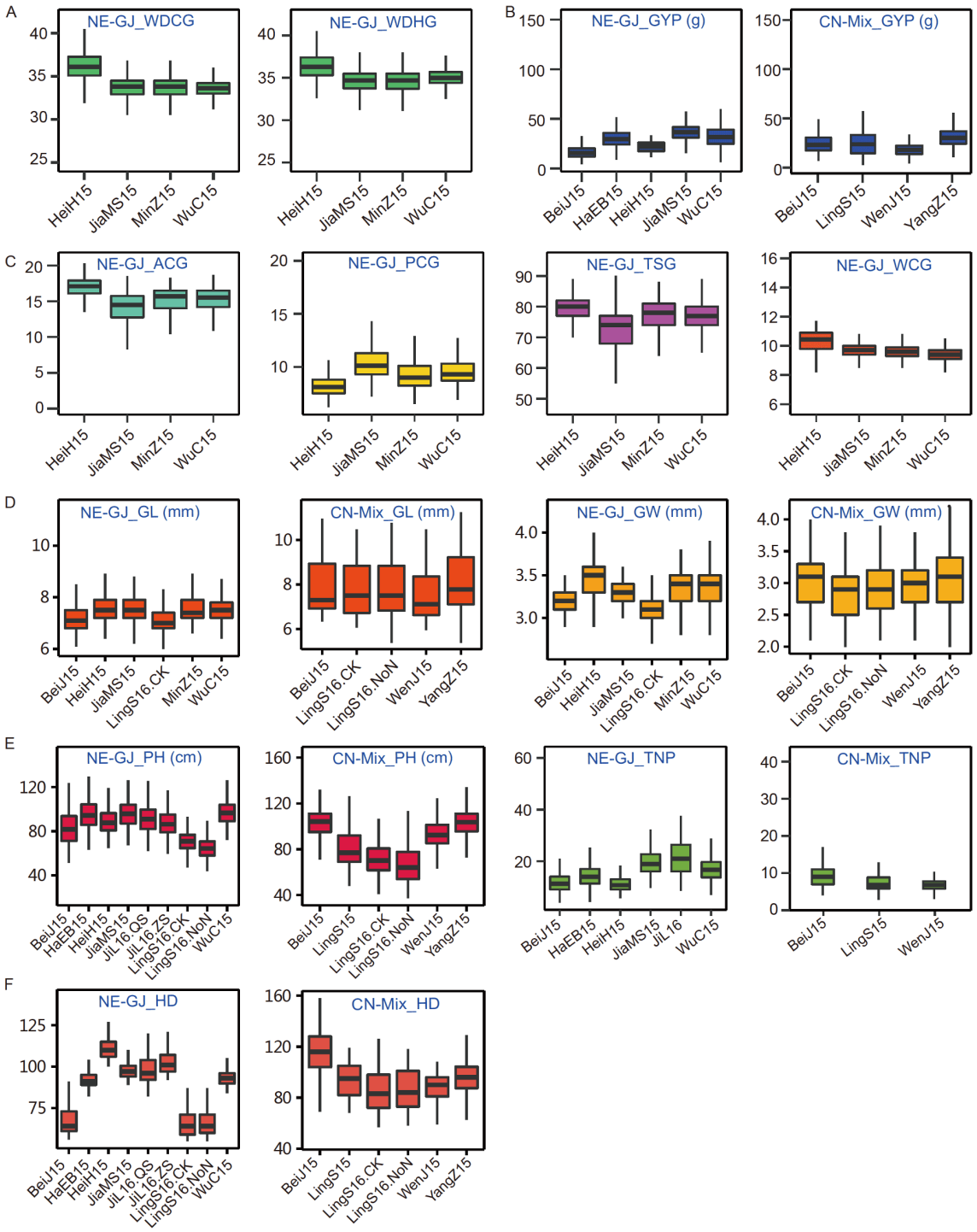

Figure 2 Multi-trait phenotyping at agro-ecologically diverse locations. A, Whiteness degree of complete grain (WDCG) and whiteness degree of dead grain (WDHG) included in grain appearance for NE-GJ population panel. B, Grain yield per plant (GYP) included in yield components for NE-GJ and CN-Mix population panels. C, Amylose content of grain (ACG), protein content of grain (PCG), taste score of cooked grain (TSG), and water content of grain (WCG) included in grain cooking characteristics for NE-GJ population panel. D, Grain length (GL) and grain width (GW) included in grain size for NE-GJ and CN-Mix population panels. E, Plant height (PH) and tiller number per plant (TNP) included in plant architecture for NE-GJ and CN-Mix population panels. F, Heading date (HD) for NE-GJ and CN-Mix population panels.

GWAS on measured traits

GWAS have revealed genetic variations associated with important agronomic traits in specific subpopulations or across several subpopulations (Huang et al., 2011; Zhao et al., 2011). On the basis of our measured phenotypic datasets,

we grouped the rice accessions into seven GWAS subpopulations: NE-GJ, CN-Mix, their subpopulations (NE-GJ1 and NE-GJ2 as two subpopulations of NE-GJ; and CN-XI and CN-GJ as two subpopulations of CN-Mix), and All-GJ for all GJ accessions. GWAS for all measured traits were performed using Efficient Mixed-Model Association Expedited, which could correct for hidden relatedness and pedigree structure (Kang et al., 2010). In total, we performed GWAS on 97 and 49 phenotypic datasets for the NE-GJ-related and CN-Mix-related panels, respectively. On the basis of the whole-genome LD decay distance of the rice population in this study and results from previously published studies (Chen et al., 2014; Crowell et al., 2016; Guo et al., 2018), we defined adjacent significant SNPs within a region of less than 200 kb as a single locus.

We detected a total of 143 significant loci in all GWAS panels (Table 2). A total of 25 traits were found to be associated with at least two significant loci (Table S7 in Supporting Information). A total of 77 loci were detected in at least two locations (Table 2). Among the GWAS signals, we found 91 loci (26 traits) that were located close to previously identified genes with known functions (Table S7 in Supporting Information). There were 94 genes in 52 GWAS regions that were detected in multiple environments, and 77 genes in 51 GWAS regions were detected in multiple subpopulations. For example, we found that several heading date genes *Hd3a*, *RFT1*, *Hd1*, *DTH7*, and *Ghd7* were repeatedly detected under different environments (Figure S3 in Supporting Information). GWAS loci around *GS3* and *GW5* showed very strong effects on grain length and width (Figures S4 and S5 in Supporting Information). Further, two other GWAS loci, *TUD1* and *OsAPC6*, also showed significant association with grain length and width in the NE-GJ panel (Figures S4 and S5 in Supporting Information). Spike length was associated with a region containing the known gene *DEP1*. Grain number was associated with the *NOG1* region. Phenotypic values were well correlated with genotypes of known genes, including *DTH7*, *Ghd7*, *Hd6*, *Hd3a*, *GS3*, *GW5*, and *NOG1*, in GWAS loci (the genotype of the leading SNP correlated with the genes) (Figure S6 in Supporting Information).

We observed 20 loci associated with multiple traits (Table S7 in Supporting Information). Some loci contain multiple genes with different functions. For example, a locus significantly associated with heading date and plant height at Chr1:36.5–40.6 Mb includes relevant known genes *OsMADS1* and a gene cluster with *sd1*, *MOC2*, *OsPdk1* and *REL1*, respectively (Figure S7A in Supporting Information). Region Chr6:27.9–28.6 Mb was observed to be significantly associated with plant height and grain width (Figure S7B in Supporting Information). This region is close to known genes *OsARF18* and *DEP3,* which control both plant height and grain size. Region Chr7:8.4–9.0 Mb was significantly associated with heading date and blighted grains per panicle (Figure S7C in Supporting Information) and contains *Ghd7*, which has pleiotropic effects on grain number, plant height, and heading date in rice (Xue et al., 2008). Region Chr7:28.8–29.7 Mb was significantly associated with heading date and plant height for NE-GJ1 panel (Figure S7D in Supporting Information) and contains *DTH7*, which has pleiotropic effects on heading date and plant height in rice (Liu et al., 2013). Further work using different populations or methods, such as sequential incorporation of different traits as a second phenotypic covariate within the mixed model, may help to determine whether the multi-trait-association loci are pleiotropic or are just gene clusters (Crowell et al., 2016).

GWAS within subpopulation-specific panels identified 42 and 46 additional association signals that were not detected in the NE-GJ and CN-Mix panels, respectively. For example, within the CN-GJ panel, we detected a GWAS locus in chromosome 1 (*Rdd1* and *OsPIL15*) that was associated with grain length (Figure S8D–F in Supporting Information). With the NE-GJ1 panel, we detected locus *Wx*, associated with grain amylose content (Figure S9A–C in Supporting Information). We identified a total of 53 significant environment-specific peaks for 18 traits (Table S7 in Supporting Information). For example, GWAS locus *Ghd7* was detected only at Lingshui for the CN-Mix panel (Figure S4 in Supporting Information). Locus *Hd3a*, associated with heading date in the CN-Mix panel, was identified at Beijing, Yangzhou, and Wenjiang, but no significant association was

Table 2 Summary of genome-wide significant associations

	Total	Population							Detected in at least two populations	Detected in at least two environments
		NE-GJ	NE-GJ1	NE-GJ2	CN-Mix	CN-XI	CN-GJ	All-GJ		
Number of traits	28	22	18	18	15	15	11	5	24	20
Number of loci[a)]	143	48	42	40	48	40	36	29	71	77
Number of new loci	93	32	28	22	30	27	23	13	46	48
Number of lead SNPs[b)]	479	96	85	55	95	66	62	40	20	32
Number of known genes[c)]	213	45	49	36	69	55	60	21	77	94

a) Adjacent significant SNPs separated by less than 200 kb were considered to be in a cluster. b) SNPs with the lowest *P* value in a defined region. c) Genes located in significant loci.

found at Lingshui (Figure S3 in Supporting Information). These results indicate the value of using multiple populations and multiple environments in GWAS.

New candidate genes or alleles identified by GWAS

With our GWAS, we found a total of 93 new loci not located close to known genes controlling the target trait, including 13 loci for six traits (ACG, PCG, HD, PH, GL, and GW) detected in at least two locations (Table S7 in Supporting Information). For example, a chromosome 1 region from 32.3 to 36.0 Mb with a lead SNP at position 34,501,250 was significantly associated with grain amylose content. On the basis of functional annotations, we identified three candidate genes, LOC_Os01g59120 (Guo et al., 2010), LOC_Os01g59660 (Kaneko et al., 2004), and LOC_Os01g61810 (Miyoshi et al., 2003), within this region. LOC_Os01g59660 (*OsGAMYB*), a MYB transcription factor that induces expression of alpha-amylase in the aleurone layer and also controls other traits (Kaneko et al., 2004), was identified as the best candidate gene on the basis of its genotypes in the population. *OsGAMYB* contains a SNP at position 34,512,532 that was highly correlated with the phenotypes and SNP Chr1P34501250 (Figure 3). SNP Chr1P34512532 includes a high risk splice acceptor site variant that was correlated with low amylose content and was only present in several early upland rice lines from the Northeast, such as Heimangdao, Wuchangbaimang, and Sanjiang1. Further work will be needed to identify downstream genes that are regulated by *OsGAMYB* to decrease the amylose content.

We also identified new alleles of known genes. Locus Chr6:2.68–4.62 Mb was found to be associated with heading date in both CN-Mix and CN-GJ panels (Figure 4A and B). This region contained *Hd3a* with two distinct functional SNPs (Chr6P2942292 and Chr6P2942293) whose variants in Kasalath and Nipponbare affected flowering under LD conditions, with Hap3 in Kasalath flowering earlier than Hap0 in Nipponbare (Figure 4D and E). The lead SNP (Chr6P2971050) in the locus correlated well with a new SNP Chr6P2942192, which corresponded with Hap2 of *Hd3a* and showed even earlier flowering than the other genotypes (Figure 4C–E). The new Hap2 allele of *Hd3a* was found in 52 total accessions, of which most were early flowering NE-GJ, and could be used as an allele donor in breeding pro-

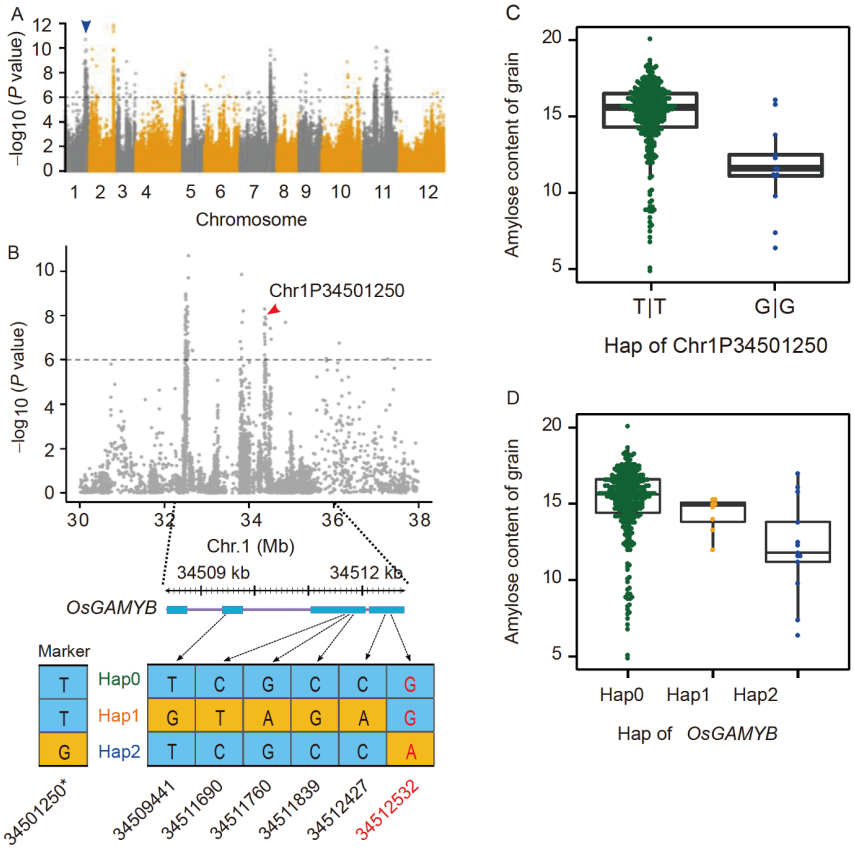

Figure 3 Association signals for amylose content with *OsGAMYB* as a candidate gene. A, Manhattan plots of GWAS results for amylose content. An arrowhead indicates the position of association loci from 32.3–36.0 Mb on chromosome 1. B, Association loci for amylose content on chromosome 1 in the NE-GJ panel. The bottom of the plot shows the region containing *OsGAMYB*. SNP Chr1P34512532 in the last exon of *OsGAMY*B is highly correlated with the association SNP at Chr1P34501250. C, Phenotypic variation of SNP Chr1P34501250. D, Phenotypic variation of *OsGAMYB* haplotypes on the basis of CDS region SNPs.

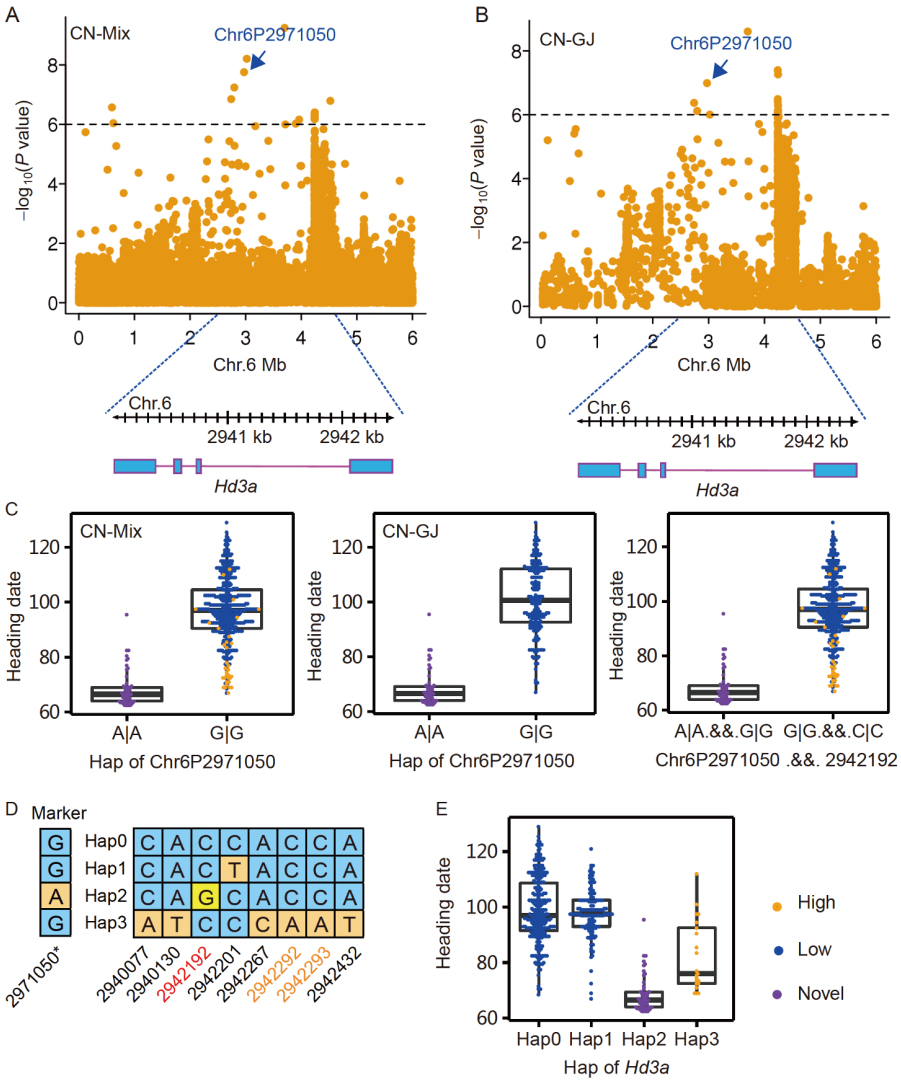

Figure 4 Association signals for heading date with identification of a novel allele of candidate gene *Hd3a*. A and B, Manhattan plots of GWAS results for heading date on chromosome 6 for CN-Mix and CN-GJ panels, respectively. The bottom of each plot depicts the *Hd3a* region. C, Phenotypic variation of association SNP Chr6P2971050. D, *Hd3a* haplotypes based on CDS region SNPs. SNP Chr6P2942192 in *Hd3a* is highly correlated with association SNP Chr6P2971050. Two non-synonymous nucleotide substitutions at Chr6P2942292 and Chr6P2942293 contributed to a new functional allele as in Kasalath (Takahashi et al., 2009). E, Phenotypic variation of *Hd3a* haplotypes based on CDS region SNPs.

grams. Another locus from 39.0–40.6 Mb on chromosome 1, which was associated with heading date in the CN-XI subpopulation, contains the gene *OsMADS51*, which is known to promote flowering under short day conditions (Figure S9A in Supporting Information). No known natural variations have been reported for this gene. We found four potential causal SNPs, including two non-synonymous-coding variants at positions 40,362,268 and 40,362,776 and two frame-shift variants at positions 40,363,925 and 40,363,928 (Figure S9B in Supporting Information). These SNPs were differentially distributed between XI and GJ accessions and strongly associated with phenotypes in both the CN-Mix and CN-XI subpopulations (Figure S9C in Supporting Information). We found two loci, Chr8:26.5–27.3Mb and Chr10:16.0–16.5 Mb, associated with grain shape. Locus Chr8:26.5–27.3 Mb included known gene *GW8*, but the alleles in our population have not been previously reported, with all of the long grain SNPs present in early flowering landraces of upland rice in the Northeast. These new loci represent candidate alleles for functional studies and breeding programs.

Agronomically favorable allele usage in super rice lines

To extract further value from our sequencing data, we selected 63 well studied genes, including 94 functionally verified natural variants (or alleles), and analyzed their distribution and functional consequences in our sequenced rice accessions (Table S8 in Supporting Information). These genes include 12 for heading date, 13 for plant architecture, nine for grain shape, six for eating or storage quality, three for fertilizer usage, eight for disease resistance, five for cold

tolerance, and several others related to male sterility, shattering, or seed dormancy. We classified the alleles into several different types: WT for wild type with normal biological function, Low for lower expression value or lower protein activity, High for higher expression value or higher protein activity, LoF for loss of gene or protein function, GoF for gain of function, and Novel for potential new functions.

Firstly, we analyzed the genetic effects of 27 agronomically important genes that have at least two alleles and corresponding phenotype data. We used multiple comparisons by least significant difference to identify significant phenotype differences among different alleles under at least one environment for one panel (Table S9 in Supporting Information). Phenotypic effects of genes were consistent with their known functions. For example, the Low and LoF alleles of *DTH7*, LoF allele of *Ghd7*, and Novel and High alleles of *Hd3a* were found to promote flowering under long day, while Low allele of *Ehd1* was found to delay flowering under short day. The LoF alleles of *GS3*, *OsMADS1*, and *GW5* were found to increase grain size for all panels and under all environments.

Secondly, we conducted direct sequence analysis to identify the usage patterns of favorable alleles in our rice population (Figure S10 in Supporting Information). We focused on a set of elite high yield potential Chinese cultivars that have been denoted as super rice by the Ministry of Agriculture and Rural Affairs of China or are parental lines of super hybrid rice cultivars (Figure 5, Table 3). For most of the 63 genes, we found that multiple allele types were used in these lines. Several favorable alleles were widely used, such as those of *SBI*, *Waxy*, *ALK*, *Rc*, *LTG1*, and *Sdr4*; however, many are not, including those for *Ef-cd*, *qNPT1*, *IPA1*, *OsMADS1*, *GW7*, *GW2*, *Badh2*, *BSR-d1*, *Pit*, *XA5*, and *CTB4a*. We also observed that many alleles are differentially used between XI and GJ. Some show no clear benefits over the other (such as *DTH2*, *SD1*, *SCM2*, *TAC1*, and *TAC3*), while others benefit either the GJ or XI population. Favorable alleles of *GW8, Phr1*, and *HAN1* are widely used in GJ lines, but not in XI lines. Favorable alleles of *Gn1a*, *NOG1*, *LAX*1, *GLW7*, *NRT1.1B*, *Pid3*, and *Pid2* are widely used in XI lines, but not in GJ lines. This allele usage information identifies potentially underused favorable alleles and candidate breeding lines that could be used as favorable allele donors in breeding programs. Further, we examined the allele usage pattern of several important agronomic traits in the elite lines.

Heading date

Heading date contributes substantially to growth duration and regional adaptation along latitude and is the most important trait accounting for grain yield. Towards optimizing growth duration in their target planting areas to gain high yield, it is clear that super rice accessions normally planted in southern regions use more late flowering alleles of major HD genes than cultivars planted in Northeast China (Figure 5). For example, the heading dates at both Beijing and Lingshui of four famous GJ cultivars—which in order from early to late were KY131, Longgeng31, Daohuaxiang2, and Ninggeng4—were consistent with their use of early or late flowering alleles of major HD genes (Figure 5). Comparisons of major HD genes, such as *Hd1*, *Ghd7*, and *DTH7*, used in GJ cultivars in southern regions and XI lines, suggest the usage of individual alleles is more likely due to their combined effect in generating a suitable heading date rather than for the biological advantage of one allele over another. For example, GJ cultivars in the South used WT alleles of *Hd1*, but the LoF alleles of *Hd1* were more frequently used in NE-GJ and the female parents of two-line hybrid XI rice and less frequently used in the female parents of three-line hybrid XI rice. XI parental lines carried alleles of major HD genes, such as *Hd3a* and *DTH8*, which can be combined with different alleles from another parent to generate hybrids with

Table 3 Summary of the allele usage for important genes in 47 super rice cultivars and the parental lines of super rice cultivars

Allele usage type	Genes
Different alleles between XI and GJ without functional advantage	*DTH2*, *SD1*, *SCM2*, *TAC1*, *TAC3*
Different between XI and GJ with favorable alleles in XI	*Gn1a*, *NOG1*, *LAX1*, *GLW7*, *NRT1.1B*, *Pid3*, *Pid2*
Different between XI and GJ with favorable alleles in GJ	*GW8*, *Phr1*, *COLD1*, *OsbZIP73*, *HAN1*
Favorable alleles preferentially present in XI	*GS3*, *Chalk5*, *Pup1*, *TOND1*
Favorable alleles preferentially present in GJ	*DEP1*, *NAL1*, *GW5*, *qSH1*
Favorable alleles in low frequency	*Ef-cd*, *qNPT1*, *IPA1*, *OsMADS1*, *GW7*, *GW2*, *Badh2*, *BSR-d1*, *Pit*, *XA5*, *CTB4a*
Favorable alleles widely used	*SBI*, *Waxy*, *ALK*, *Rc*, *LTG1*, *Sdr4*
No functional advantage with minor alleles in XI	*RFT1*, *Hd3a*
No functional advantage with minor alleles in GJ	*Hd6*, *Hd18*, *OsTB1*
Minor alleles for special use	*TMS5*, *PMS3*
Multiple alleles without functional advantage	*Hd1*, *DTH7*, *Ghd7*, *Hd17*
Multiple favorable alleles without preference	*STV11*, *Pi-ta*, *Ptr*

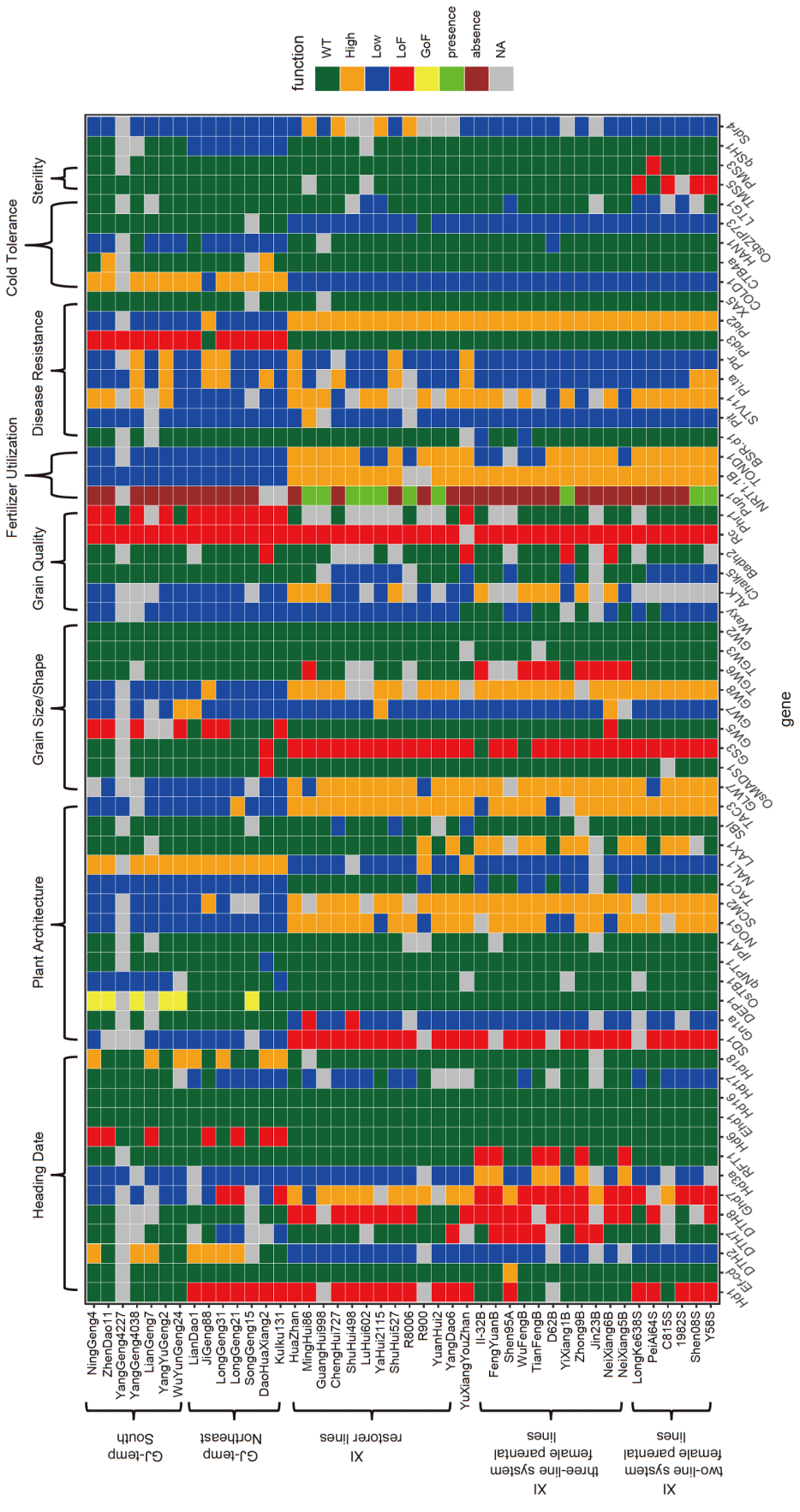

Figure 5 Allele types of 63 important genes in super rice cultivars or the parental lines of super hybrid rice cultivars. Allele types are represented by different colors: WT, wild type with normal biological function; Low, low RNA expression value or protein activity; High, high RNA expression value or protein activity; LoF, loss of gene or protein function; GoF, gain of new gene function.

improved hybrid vigor and optimized growth periods (Huang et al., 2016; Huang et al., 2015; Li et al., 2016).

The HD genes allele usage pattern suggests that changing specific alleles may have helped elite cultivars to adapt to new growing areas. For example, adding the *DTH7* and *Ghd7* early flowering alleles to Daohuaxiang2 may have helped it grow further north. Conversely, the introduction of the *DTH7* and *Ghd7* late flowering alleles to KY131 may have helped it grow better in southern regions. Notably, a favorable allele of the recently cloned *Ef-cd* gene can reduce growth duration without sacrificing yield (Fang et al., 2019); however, it is currently used in only one of the super rice lines, Shen95A. Therefore, expanded use of the *Ef-cd* early flowering allele has a strong potential for improving XI or GJ lines in the South. However, to optimize the growth period to maximize a cultivar's grain yield in the same planting region, some other later flowering alleles may also be needed in combination with the *Ef-cd* early flowering allele.

Plant architecture

Plant architecture is one of the major targets of current rice breeding. Several genes are differentially used between GJ and XI lines, such as *NOG1*, *SCM2*, *TAC1*, *NAL1*, and *TAC3*, reflecting the major phenotypic difference between the two subspecies. Many super rice lines, such as Ninggeng4, Longgeng31, and Y58S, also use the favorable alleles of genes controlling architecture, such as GoF of *DEP1* and WT or High of *OsTB1*. The favorable allele of ideal architecture gene *qNPT1* was found only in Daohuaxiang2. It is worth mentioning that WuFengB uses the *TAC1* Low allele in combination with the *TAC1* High allele from paternal lines to generate hybrid vigor (Yu et al., 2007).

The allele usage in super rice lines suggests that a strategy for potential improvement is to combine favorable alleles, such as *Gn1a* Low and *NOG1* High allele, to increase the grain number per spike, especially in GJ cultivars. Further, the introduction of XI alleles of *SCM2*, for example, to Daohuaxiang2 may reduce its high lodging-susceptibility, a key breeding target for its further improvement. Underused favorable alleles of *IPA1* and *qNPT1* in XI also have great potential for use in future breeding.

Grain shape and quality

Grain shape is the most stable trait under varying environments, as shown in our multi-environmental data, and thus is probably the easiest target to manipulate in breeding by selection of desired alleles of known genes. The allele usage of major genes G*W5* and *GS3*, as well as several other minor genes, is consistent with GJ cultivars mostly having short and round grain, while XI lines mostly have long grain. Daohuaxiang2, different from most of the temperate GJ cultivars, carries *GS3* LoF, *GW5* wild, and *OsMADS1* LoF, resulting in long grain similar to XI rice. Favorable alleles of *OsMADS1*, *TGW3*, and *GW2* increased the grain size without obvious side effects but were present in only a small number of accessions, indicating great potential for its use in future breeding.

For grain quality, the *Waxy* and *ALK* low allele produce superior cooking and eating qualities in most of the super lines. The lack of the favorable alleles of *Waxy*, *ALK*, or *Chalk5* in several XI cultivars, such as Huazhan and WufengB, suggests their potential use in future XI breeding. GJ cultivars may also benefit from the addition of the favorable allele of *Chalk5*. Daohuaxiang2 is an excellent donor for the *BADH2* LoF allele, which improves grain fragrance. Compared with many other GJ super rice cultivars containing the same alleles of known genes, KY131 has a higher eating quality, suggesting the existence of unknown genes that contribute to the high eating quality of KY131.

Other traits

XI lines in general have better blast resistance than GJ accessions. For example, KY131 carries the sensitive alleles of most major blast resistance genes, which is consistent with its high sensitivity to rice blast. GJ cultivars, especially KY131, would benefit from the addition of resistance alleles of broad-spectrum disease resistance genes, such as *Pigm*, *Bsr-d1*, *Xa5* and *Pit*. Cold tolerance is important for the normal growth and high grain yield of NE-GJ. Famous GJ cultivars, such as KY131, Longgeng31, and Daohuaxiang2, all carry cold tolerance alleles of *COLD1*, *OsbZIP73*, and *qPSR10*. Daohuaxiang2 also carries the *CTB4a* cold tolerance allele, but its cold tolerance is weaker than that of KY131, suggesting the existence of unknown cold tolerance genes in KY131. High fertilizer usage efficiency is becoming more and more important for reducing fertilizer usage to protect environments. Most XI rice lines, but fewer GJ lines, used favorable alleles of *NRT1.1B* and *TOND1* for high N-usage efficiency.

In summary, we have found that many favorable alleles have been underused in either GJ lines, XI lines, or in both populations, and many elite lines could be further improved by adding unused favorable alleles. In addition, we have found that KY131 may be an excellent source for further improvement of Northeast cultivars because it contains unidentified genes that could be important for breeding.

DISCUSSION

In this study, we report large-scale genomic and multi-environmental phenomic datasets of Chinese rice cultivars and elite parental lines. Unlike many previously studied rice populations (Crowell et al., 2016; Huang et al., 2010; Huang et al., 2011; Wang et al., 2018b; Yang et al., 2014), the majority of rice accessions we studied here have undergone

intensive modern breeding improvements. Therefore, these accessions contain many agronomically important alleles that can be repeatedly used in future breeding programs and should be functionally studied.

We measured multiple agronomic traits at several agro-ecologically diverse locations. Most agronomically important traits are influenced by many genetic and non-genetic factors, which indicates the need for detailed phenotyping of rice accessions at diverse locations under different environments and treatments (Huang et al., 2015). We detected a total of 143 GWAS association signals, including many subpopulation-specific or environment-specific trait associations and several signals associated with multiple traits, confirming the benefit of using large-scale populations and multi-environmental phenotyping. Many association loci detected here, such as *DTH7*, *Hd1*, *Ghd7*, *Gn1a*, and *DEP1*, have been widely used for rice improvement over the long history of Chinese rice breeding, and they will continue to offer improvements in future rice breeding (Feng et al., 2017; Gao et al., 2014; Wang et al., 2015a; Wang et al., 2018a; Xu et al., 2016; Ye et al., 2018; Zhang et al., 2015).

Molecular breeding by rational design is a new breeding strategy for quick and targeted breeding improvement that combines known favorable gene alleles and QTLs (Guo et al., 2019). For example, elite cultivars with high yield and quality were bred by pyramiding multiple favorable alleles (Zeng et al., 2017). Direct sequence analysis of known genes is complementary to GWAS in that it is difficult and unnecessary to use a population with measured phenotypes to detect known genes. By direct sequence analysis, we found that many favorable alleles remain underused in elite cultivars, suggesting that some elite lines could be improved by rational design by adding unused favorable alleles. In recent years, substantial ongoing breeding efforts have been conducted to improve some of these lines. For example, several important KY131 traits, including blast resistance, grain length, and grain number, have been improved, leading to new elite cultivars (Feng et al., 2019; Feng et al., 2017; Nan et al., 2018; Wang et al., 2019). In summary, our large-scale genotyping and phenotyping datasets provide a valuable resource for future molecular genetic studies of Chinese rice and rice improvement.

METHODS

Sampling, DNA sample preparation, and sequencing

O. sativa accessions were collected from breeders and planted in Beijing for whole-genome resequencing. Total genomic DNA was isolated from young leaf tissue of each plant. Paired-end sequencing libraries with insert sizes of 450±50 bp were constructed for each sample in accordance with Illumina's standard protocols and then sequenced as 125–150 bp paired-end reads on Illumina Hi-Seq sequencing systems (Illumina, USA). Raw sequences were further processed to remove adaptors and low-quality reads, yielding an average of around 2.5 Gb for each sample.

Sequence variation calling

Paired-end reads of all varieties were aligned to the rice reference genome (Nipponbare, MSU version 7.0) using the MEM algorithm of BWA (Li and Durbin, 2009). SAMtools (version 1.7) (Li, 2011) was used to sort BAM files resulting from BWA output and remove duplicates to keep reads with a mapping quality above Q30. The sorted and filtered alignment files were then input into Genome Analysis Toolkit (GATK, V3.4-46) (McKenna et al., 2010) to call SNPs. Multiple SNP calling was performed using the UnifiedGenotyper of GATK. SNPs assigned as 'PASS' by GATK and having a missing rate less than 0.2 and a minor allele frequency greater than 0.05 were retained for further analysis.

Population genetics analysis

A neighbor-joining tree, principal component analysis, and structure plots were used to infer the population structure of the *O. sativa* accessions. The neighbor-joining tree was constructed using PHYLIP software (version 3.697) (http://evolution.genetics.washington.edu/phylip.html) on the basis of a pairwise distance matrix derived from the simple matching distance for all CDS-region SNPs. The software FigTree (http://tree.bio.ed.ac.uk/software/figtree/) was used to visualize the phylogenetic tree. Principal component analysis was performed using EIGENSOFT software (version 6.0.1) (Price et al., 2006). To minimize the contribution from extensive strong LD regions, the SNPs included in PCA analysis were pruned for LD using PLINK (-indep-pairwise 50 5 r2) (http://pngu.mgh.harvard.edu/purcell/plink/). The first three principal components were used for PCA plots. The maximum-likelihood clustering program ADMIXTURE (version 1.23) (Alexander et al., 2009) was used to estimate the optimum number of clusters for the 1,275 rice accessions. Whole genome LD was estimated using pairwise r^2 between SNPs with parameters -ld-window-r2 0 -ld-window 99999 -ld-window-kb 1000 in PLINK (http://pngu.mgh.harvard.edu/purcell/plink/). The r^2 value was calculated within a corresponding chromosome, and then pairwise r^2 values were averaged across the whole genome for LD decay estimation.

Phenotyping

Rice samples were planted for at least one season in multiple locations for different types of trait phenotyping. Rice sam-

ples in the NE-GJ population panel were planted at eight ecologically diverse locations: Beijing, Jilin, Lingshui, and five cities in four accumulated temperature zones in Heilongjiang province: Haerbin, Minzhu, Heihe, Wuchang, and Jiamusi. CN-Mix samples were planted at four diverse locations: Beijing, Yangzhou, Wenjiang, and Lingshui.

The phenotyping in this study involved a wide range of agronomic traits for yield components. Field traits, including heading date, plant height, tiller number, panicle number, panicle length, flag leaf length, flag leaf width, and the flag leaf soil-plant analysis development (SPAD) value (a parameter indicating the relative chlorophyll content measured by SPAD502), were measured directly in the field. Heading date was defined as the number of days from sowing to inflorescence emergence above the flag leaf sheath for just more than half of the individuals. The other field traits were measured for at least three samples of each accession with standard methods, and their average was used. The yield and grain related traits, including grain yield per plant, grain number per panicle, grain length, grain width, and grain weight per 1,000 grains, were measured in the laboratory after harvest. Among them, grain length, grain width, and grain weight were measured with a Wanshen SC-G automatic grain test instrument (Wanshen SC-G, China). Fully filled grains with low grain moisture content were used to measure grain quality traits with a grain appearance analyzer (Foss Infratec IM 1241) and Wanshen SC-E rice appearance quality detector (Wanshen SC-E, China).

Genome-wide association study (GWAS)

GWAS was independently performed with multiple panels. Only SNPs from each GWAS panel with a minor allele frequency greater than 0.05 and missing rate less than 0.2 were used for GWAS. GWAS using a mixed model was performed with EMMAX software (Kang et al., 2010). Population stratification and relatedness were modeled with a kinship (K) matrix in the emmax-kin-intel package of EMMAX. The K matrix was used as a random effect and the first three principal components were included as fixed effects. For all six GWAS panels, SNPs with a genome-wide *P*-value less than 1.0×10^{-6} were denoted as significant association SNPs.

Genotype analysis of known genes

Causal mutations have been reported for many known genes, which were called functionally verified natural variations. For most of these genes, Nipponbare was used as the reference to identify their variations in the population; however, a few genes, such as *GW5* and *Pup1*, were either absent or contained large sequence deletions in Nipponbare, so R498 (Du et al., 2017) was used as the reference for analyzing their allele types. Rice accession genotypes were identified using the following method. SNPs and small indels were identified using GATK unifiedGenotyper; large indels relative to Nipponbare were identified on the basis of sequence coverage; and genes that were absent in Nipponbare but present in R498 were genotyped similarly. Genotypes that could not be unambiguously identified with resequencing data were denoted as N/A. For only a limited number of disease resistance genes, allele types were unambiguously identified due to the relative low sequencing depth and the complexity of their sequences.

Accession numbers

All raw reads generated for rice accessions used in this study have been deposited in the National Genomics Data Center with BioProject PRJCA000322 and GSA accession CRA000167.

Compliance and ethics *The author(s) declare that they have no conflict of interest.*

Acknowledgements *This work was partially supported by the Chinese Academy of Sciences "Strategic Priority Research Program" fund (XDA08020302) and grants from State Key Laboratory of Plant Genomics.*

References

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

Bai, W., Zhang, H., Zhang, Z., Teng, F., Wang, L., Tao, Y., and Zheng, Y. (2010). The evidence for non-additive effect as the main genetic component of plant height and ear height in maize using introgression line populations. Plant Breed 129, 376–384.

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.

Crowell, S., Korniliev, P., Falcão, A., Ismail, A., Gregorio, G., Mezey, J., and McCouch, S. (2016). Genome-wide association and high-resolution phenotyping link *Oryza sativa* panicle traits to numerous trait-specific QTL clusters. Nat Commun 7, 10527.

Doi, K., Izawa, T., Fuse, T., Yamanouchi, U., Kubo, T., Shimatani, Z., Yano, M., and Yoshimura, A. (2004). *Ehd1*, a B-type response regulator in rice, confers short-day promotion of flowering and controls FT-like gene expression independently of *Hd1*. Genes Dev 18, 926–936.

Dong, H., Zhao, H., Li, S., Han, Z., Hu, G., Liu, C., Yang, G., Wang, G., Xie, W., and Xing, Y. (2018). Genome-wide association studies reveal that members of *bHLH* subfamily 16 share a conserved function in regulating flag leaf angle in rice (*Oryza sativa*). PLoS Genet 14, e1007323.

Du, H., Yu, Y., Ma, Y., Gao, Q., Cao, Y., Chen, Z., Ma, B., Qi, M., Li, Y., Zhao, X., et al. (2017). Sequencing and *de novo* assembly of a near complete *indica* rice genome. Nat Commun 8, 15324.

Duan, P., Xu, J., Zeng, D., Zhang, B., Geng, M., Zhang, G., Huang, K., Huang, L., Xu, R., Ge, S., et al. (2017). Natural variation in the promoter of *GSE5* contributes to grain size diversity in rice. Mol Plant 10, 685–694.

Fang, J., Zhang, F., Wang, H., Wang, W., Zhao, F., Li, Z., Sun, C., Chen, F., Xu, F., Chang, S., et al. (2019). *Ef-cd* locus shortens rice maturity

duration without yield penalty. Proc Natl Acad Sci USA 116, 18717–18722.
Feng, X., Lin, K., Zhang, W., Nan, J., Zhang, X., Wang, C., Wang, R., Jiang, G., Yuan, Q., and Lin, S. (2019). Improving the blast resistance of the elite rice variety Kongyu-131 by updating the *pi21* locus. BMC Plant Biol 19, 249.
Feng, X., Wang, C., Nan, J., Zhang, X., Wang, R., Jiang, G., Yuan, Q., and Lin, S. (2017). Updating the elite rice variety Kongyu 131 by improving the *Gn1a* locus. Rice 10, 35.
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 USA 111, 16337–16342.
Guo, J., Wang, F., Song, J., Sun, W., and Zhang, X.S. (2010). The expression of Orysa;*CycB1;1* is essential for endosperm formation and causes embryo enlargement in rice. Planta 231, 293–303.
Guo, T., Yu, H., Qiu, J., Li, J., Han, B., and Lin, H. (2019). Advances in rice genetics and breeding by molecular design in China (in Chinese). Sci Sin Vitae 49, 1185–1212.
Guo, Z., Yang, W., Chang, Y., Ma, X., Tu, H., Xiong, F., Jiang, N., Feng, H., Huang, C., Yang, P., et al. (2018). Genome-wide association studies of image traits reveal genetic architecture of drought resistance in rice. Mol Plant 11, 789–805.
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.
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., Yang, S., Gong, J., Zhao, Q., Feng, Q., Zhan, Q., Zhao, Y., Li, W., Cheng, B., Xia, J., et al. (2016). Genomic architecture of heterosis for yield traits in rice. Nature 537, 629–633.
Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., Li, W., Zhan, Q., Cheng, B., Xia, J., et al. (2015). Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. Nat Commun 6, 6258.
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.
Kaneko, M., Inukai, Y., Ueguchi-Tanaka, M., Itoh, H., Izawa, T., Kobayashi, Y., Hattori, T., Miyao, A., Hirochika, H., Ashikari, M., et al. (2004). Loss-of-function mutations of the rice *GAMYB* gene impair α-amylase expression in aleurone and flower development. Plant Cell 16, 33–44.
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.
Kovi, M.R., Sablok, G., Bai, X.F., Wendell, M., Rognli, O.A., Yu, H.H., and Xing, Y.Z. (2013). Expression patterns of photoperiod and temperature regulated heading date genes in *Oryza sativa*. Comput Biol Chem 45, 36–41.
Li, D., Huang, Z., Song, S., Xin, Y., Mao, D., Lv, Q., Zhou, M., Tian, D., Tang, M., Wu, Q., et al. (2016). Integrated analysis of phenome, genome, and transcriptome of hybrid rice uncovered multiple heterosis-related loci for yield increase. Proc Natl Acad Sci USA 113, E6026–E6035.
Li, G., Jin, J., Zhou, Y., Bai, X., Mao, D., Tan, C., Wang, G., and Ouyang, Y. (2019). Genome-wide dissection of segregation distortion using multiple inter-subspecific crosses in rice. Sci China Life Sci 62, 507–516.
Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27, 2987–2993.
Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Liu, T., Liu, H., Zhang, H., and Xing, Y. (2013). Validation and characterization of *Ghd7.1*, a major quantitative trait locus with pleiotropic effects on spikelets per panicle, plant height, and heading date in rice (*Oryza sativa* L.). J Integr Plant Biol 55, 917–927.
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.
Miyoshi, K., Ito, Y., Serizawa, A., and Kurata, N. (2003). *OsHAP3* genes regulate chloroplast biogenesis in rice. Plant J 36, 532–540.
Nan, J., Feng, X., Wang, C., Zhang, X., Wang, R., Liu, J., Yuan, Q., Jiang, G., and Lin, S. (2018). Improving rice grain length through updating the *GS3* locus of an elite variety Kongyu 131. Rice 11, 21.
Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904–909.
Song, Y.L., Gao, Z.C., and Luan, W.J. (2012). Interaction between temperature and photoperiod in regulation of flowering time in rice. Sci China Life Sci 55, 241–249.
Takahashi, Y., Teshima, K.M., Yokoi, S., Innan, H., and Shimamoto, K. (2009). Variations in *Hd1* proteins, *Hd3a* promoters, and *Ehd1* expression levels contribute to diversity of flowering time in cultivated rice. Proc Natl Acad Sci USA 106, 4555–4560.
Wang, J., Xu, H., Li, N., Fan, F., Wang, L., Zhu, Y., and Li, S. (2015a). Artificial selection of *Gn1a* plays an important role in improving rice yields across different ecological regions. Rice 8, 37.
Wang, Q., Xie, W., Xing, H., Yan, J., Meng, X., Li, X., Fu, X., Xu, J., Lian, X., Yu, S., et al. (2015b). Genetic architecture of natural variation in rice chlorophyll content revealed by a genome-wide association study. Mol Plant 8, 946–957.
Wang, R., Jiang, G., Feng, X., Nan, J., Zhang, X., Yuan, Q., and Lin, S. (2019). Updating the genome of the elite rice variety Kongyu131 to expand its ecological adaptation region. Front Plant Sci 10, 288.
Wang, S., Ma, B., Gao, Q., Jiang, G., Zhou, L., Tu, B., Qin, P., Tan, X., Liu, P., Kang, Y., et al. (2018a). Dissecting the genetic basis of heavy panicle hybrid rice uncovered *Gn1a* and *GS3* as key genes. Theor Appl Genet 131, 1391–1403.
Wang, W., Mauleon, R., Hu, Z., Chebotarov, D., Tai, S., Wu, Z., Li, M., Zheng, T., Fuentes, R.R., Zhang, F., et al. (2018b). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557, 43–49.
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 USA 112, E5411–E5419.
Xing, Y., and Zhang, Q. (2010). Genetic and molecular bases of rice yield. Annu Rev Plant Biol 61, 421–442.
Xu, H., Zhao, M., Zhang, Q., Xu, Z., and Xu, Q. (2016). The *DENSE AND ERECT PANICLE 1* (*DEP1*) gene offering the potential in the breeding of high-yielding rice. Breed Sci 66, 659–667.
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, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., Fang, W., Feng, H., Xie, W., Lian, X., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat Commun 5, 5087.
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, 2473–2483.
Ye, J., Niu, X., Yang, Y., Wang, S., Xu, Q., Yuan, X., Yu, H., Wang, Y., Wang, S., Feng, Y., et al. (2018). Divergent *Hd1*, *Ghd7*, and *DTH7*

alleles control heading date and yield potential of *Japonica* rice in Northeast China. Front Plant Sci 9, 35.
Yu, B., Lin, Z., Li, H., Li, X., Li, J., Wang, Y., Zhang, X., Zhu, Z., Zhai, W., Wang, X., et al. (2007). *TAC1*, a major quantitative trait locus controlling tiller angle in rice. Plant J 52, 891–898.
Zeng, D., Tian, Z., Rao, Y., Dong, G., Yang, Y., Huang, L., Leng, Y., Xu, J., Sun, C., Zhang, G., et al. (2017). Rational design of high-yield and superior-quality rice. Nat Plants 3, 17031.
Zhang, J., Zhou, X., Yan, W., Zhang, Z., Lu, L., Han, Z., Zhao, H., Liu, H., Song, P., Hu, Y., et al. (2015). Combinations of the *Ghd7*, *Ghd8* and *Hd1* genes largely define the ecogeographical adaptation and yield potential of cultivated rice. New Phytol 208, 1056–1066.
Zhang, Y., Li, Y., Wang, Y., Liu, Z., Liu, C., Peng, B., Tan, W., Wang, D., Shi, Y., Sun, B., et al. (2010). Stability of QTL across environments and QTL-by-environment interactions for plant and ear height in maize. Agric Sci China 9, 1400–1412.
Zhao, K., Tung, C.W., Eizenga, G.C., Wright, M.H., Ali, M.L., Price, A.H., Norton, G.J., Islam, M.R., Reynolds, A., Mezey, J., et al. (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2, 467.
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.

SUPPORTING INFORMATION

Figure S1 SNP density and distribution across the genome.

Figure S2 Linkage disequilibrium differentiation in different rice varietal groups.

Figure S3 GWAS of heading date for NE-GJ and CN-Mix GWAS panels.

Figure S4 GWAS of grain length for NE-GJ and CN-Mix GWAS panels.

Figure S5 GWAS of grain width for NE-GJ and CN-Mix GWAS panels.

Figure S6 The phenotype values were correlated well with genotypes of agronomically important functional genes.

Figure S7 Signals associated with multiple traits.

Figure S8 Local Manhattan plots for amylase content and grain length.

Figure S9 Association signals for heading date with *OsMADS51* as a candidate gene.

Figure S10 The allele types of 63 genes containing functionally verified natural variants in all 1,275 rice accessions.

Table S1 The list of collected 1,275 rice accessions as well as their subpopulation classification and sequence information

Table S2 List of phenotypes used for GWAS

Table S3 The agro-ecologically diverse locations of multiple agronomic traits collected for NE-GJ and CN-Mix population panels

Table S4 The agronomic traits and multi-environmental phenotypes used in three NE-GJ-related GWAS panels

Table S5 The agronomic traits and multi-environmental phenotypes used in three CN-Mix-related GWAS panels

Table S6 The Pearson's correlation between different locations in pair for all measured traits in multi-environments

Table S7 The associated loci and known genes located closely of all GWAS panels

Table S8 The allele types of 63 genes containing functionally verified natural variations in all 1,275 rice accessions

Table S9 The multiple comparisons (LSD) were evaluated for all known alleles of 27 agronomically important genes

The supporting information is available online at http://life.scichina.com and https://link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.