



Genome-Wide Association Study of Seed Dormancy and the Genomic Consequences of Improvement Footprints in Rice (*Oryza sativa* L.)

Qing Lu^{1,2†}, Xiaojun Niu^{††}, Mengchen Zhang¹, Caihong Wang¹, Qun Xu¹, Yue Feng¹, Yaolong Yang¹, Shan Wang¹, Xiaoping Yuan¹, Hanyong Yu¹, Yiping Wang¹, Xiaoping Chen², Xuanqiang Liang² and Xinghua Wei^{1*}

¹ State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou, China, ² Crops Research Institute, Guangdong Academy of Agricultural Sciences, South China Peanut Sub-Center of National Center of Oilseed Crops Improvement and Guangdong Provincial Key Laboratory of Crop Genetic Improvement, Guangzhou, China

OPEN ACCESS

Edited by:

Giampiero Valè,
Cereal Research Centre, CRA-CER,
Italy

Reviewed by:

Robert Henry,
The University of Queensland,
Australia
Alberto Gianinetti,
Centro di Ricerca Genomica e
Bioinformatica (CREA-GB), Italy
Tian-Qing Zheng,
Institute of Crop Sciences (CAAS),
China

*Correspondence:

Xinghua Wei
weixinghua@caas.cn

[†]These authors have contributed
equally to this work.

Specialty section:

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

Received: 17 October 2017

Accepted: 18 December 2017

Published: 05 January 2018

Citation:

Lu Q, Niu X, Zhang M, Wang C, Xu Q,
Feng Y, Yang Y, Wang S, Yuan X,
Yu H, Wang Y, Chen X, Liang X and
Wei X (2018) Genome-Wide
Association Study of Seed Dormancy
and the Genomic Consequences of
Improvement Footprints in Rice (*Oryza*
sativa L.). *Front. Plant Sci.* 8:2213.
doi: 10.3389/fpls.2017.02213

Seed dormancy is an important agronomic trait affecting grain yield and quality because of pre-harvest germination and is influenced by both environmental and genetic factors. However, our knowledge of the factors controlling seed dormancy remains limited. To better reveal the molecular mechanism underlying this trait, a genome-wide association study was conducted in an *indica*-only population consisting of 453 accessions genotyped using 5,291 SNPs. Nine known and new significant SNPs were identified on eight chromosomes. These lead SNPs explained 34.9% of the phenotypic variation, and four of them were designed as dCAPS markers in the hope of accelerating molecular breeding. Moreover, a total of 212 candidate genes was predicted and eight candidate genes showed plant tissue-specific expression in expression profile data from different public bioinformatics databases. In particular, LOC_Os03g10110, which had a maize homolog involved in embryo development, was identified as a candidate regulator for further biological function investigations. Additionally, a polymorphism information content ratio method was used to screen improvement footprints and 27 selective sweeps were identified, most of which harbored domestication-related genes. Further studies suggested that three significant SNPs were adjacent to the candidate selection signals, supporting the accuracy of our genome-wide association study (GWAS) results. These findings show that genome-wide screening for selective sweeps can be used to identify new improvement-related DNA regions, although the phenotypes are unknown. This study enhances our knowledge of the genetic variation in seed dormancy, and the new dormancy-associated SNPs will provide real benefits in molecular breeding.

Keywords: genome-wide association study, seed dormancy, selective sweeps, improvement footprints, rice (*Oryza sativa* L.)

INTRODUCTION

Plant seeds, especially cereal grains, are vitally important sources of human nutrition, providing half of the global per capita energy intake. Consequently, various seed traits such as seed dormancy, which is regarded as the failure of an intact viable seed to complete germination under favorable conditions (Bewley, 1997), have been under strong artificial and natural selection during crop

domestication (Kovach et al., 2007; Izawa et al., 2009). Because seed dormancy is closely related to pre-harvest sprouting (PHS), it is one of the most important traits in rice breeding programs (Bewley and Black, 1982). Rice seed dormancy is like a double-edged sword in terms of cultivation and utilization. Weak dormancy leads to a higher PHS rate in rainy weather and results in production losses and poor quality. In southern China, because of the long rainy season, it causes heavy PHS of 5–20% of hybrid rice grains (Hu et al., 2003). However, deep seed dormancy, especially for hybrid seeds, can cause non-uniform germination or even prevent germination in the process of sowing. Thus, balancing the advantages and disadvantages of seed dormancy is an important goal in rice breeding. Moreover, understanding the genetic variation of seed dormancy is of great interest to plant breeders.

The expression of seed dormancy is influenced by environmental factors such as temperature and humidity (Basbouss-Serhal et al., 2016), endogenous hormones such as abscisic acid (ABA), gibberellic acid (GA), and the ABA-to-GA ratio (White and Rivin, 2000; Fang et al., 2008; Zhang H. et al., 2009; Gu et al., 2011; Ye et al., 2015) and some special biological tissues such as the embryo, endosperm and maternal tissues (e.g., the seed coat) (Gu et al., 2015). Recently, many quantitative trait loci (QTLs) affecting seed dormancy or germination-related traits have been identified in plant species such as barley (Ullrich et al., 1993; Gao et al., 2003; Li et al., 2004), wheat (Flintham et al., 2002; Kulwal et al., 2010; Liu et al., 2013), rice (Cai and Morishima, 2000; Guo et al., 2004; Wan et al., 2006; Gu et al., 2010; Sugimoto et al., 2010; Xie et al., 2011; Lee et al., 2017), oats (Fennimore et al., 1999), sorghum (Lijavetzky et al., 2000), *Arabidopsis* (Alonso-Blanco et al., 2003; Amiguet-Vercher et al., 2015), sunflower (Gandhi et al., 2005), rye (Masojć et al., 2007, 2009), rapeseed (Feng et al., 2009; Schatzki et al., 2013), and lettuce (Huo et al., 2013). However, only a few QTLs have been finely mapped and cloned in rice, such as *qSD12* (Gu et al., 2010) on chromosome 12, *Sdr4* (Sugimoto et al., 2010) and *qSD7-1/Rc* (Gu et al., 2011) on chromosome 7 and *qSD1-2* (Ye et al., 2015) on chromosome 1.

Plant seed dormancy is a complex agronomic trait. Traditional bi-parental QTL mapping is limited by the recombination events occurring over a few generations during the development of a recombination inbred line population. Therefore, most of the biological mechanisms of seed dormancy have not been clearly elucidated. In previous reports, only three genes, *qSD12* (Gu et al., 2010), *qSD7-1* (Gu et al., 2011), and *qSD1-2* (Ye et al., 2015), were identified to regulate hormone accumulation in developing or

mature seeds in rice. Thus, more novel QTLs for this trait need to be isolated by highly efficient and reliable QTL mapping methods in the future. Genome-wide association studies (GWASs) based on the historic recombination in a large natural population, a high-density SNP map and a comprehensive HapMap have become a powerful complementary approach for linkage mapping to identify complex trait variation at the genome-wide level (Huang et al., 2010). GWASs can overcome the limitations of traditional bi-parental populations and dissect complex traits with high mapping resolution. In the past few years, GWASs have been successfully applied in the dissection of complex traits in humans (Edwards et al., 2005), animals (Duijvesteijn et al., 2010), and plant species such as *Arabidopsis thaliana* (Atwell et al., 2010), soybean (Wen et al., 2014; Zhang et al., 2015), wheat (Sukumaran et al., 2015), barley (Gawenda et al., 2015), maize (Zhang et al., 2014), sorghum (Zhang D. et al., 2015), tomato (Zhang et al., 2016a), rapeseed (Qu et al., 2015), sesame (Wei et al., 2015), and *Aegilops tauschii* (Liu et al., 2015). In rice, GWASs have also been applied widely to detect novel QTLs involved in different complex traits, for example yield (Begum et al., 2015; Huang X. et al., 2015), environmental stress resistance (Kumar et al., 2015; Lv et al., 2015), grain quality (Huang Y. et al., 2015), blast resistance (Wang C. et al., 2014; Wang et al., 2015), flowering time (Huang et al., 2011) and also seed dormancy (Magwa et al., 2016). Nowadays, together with linkage analysis, the GWAS strategy is playing an increasingly important role in dissecting novel QTLs and uncovering whole genome-wide variation.

Rice domestication has been a hot research topic for a long time. Many domestication related-genes have been mapped or cloned in previous studies, such as *PROG1* (Jin et al., 2008; Tan et al., 2008), *An1* (Luo et al., 2013), *sh4* (Li et al., 2006), *Bh4* (Zhu et al., 2011), and *sd1* (Asano et al., 2011). Seed dormancy is a typical and direct domestication-related trait that has been strongly affected by nature and human selection over the long history of rice domestication. Two seed dormancy genes, *Sdr4* (Sugimoto et al., 2010) and *qSD7-1/Rc* (Gu et al., 2011), have been proven to be directly involved in rice domestication. Long-term domestication has dramatically changed phenotypes, reduced allele frequencies, genetic diversity and polymorphism information content (PIC), and drastically increased linkage disequilibrium (LD) (Richards et al., 2004; Chen et al., 2010; Qanbari et al., 2014). Many methods based on the high density of single nucleotide polymorphism (SNP) markers have been used to detect genomic selection signals, such as nucleotide diversity (π) (Huang et al., 2012), PIC (Rostoks et al., 2006), population differentiation (F_{ST}) (Wilkinson et al., 2013), cross-population composite likelihood ratio (XP-CLR) (Xu et al., 2011) and extended haplotype homozygosity (EHH) (Sabeti et al., 2002; Olsen et al., 2006). In the case of rice, multiple selective sweeps have been detected using these methods (Olsen et al., 2006; Huang et al., 2012; Huang X. et al., 2015). Therefore, selective sweep analysis is another approach to identify QTLs for domestication traits.

In this study, on the basis of our previous reports (Lu et al., 2015), a high-density custom-designed array containing 5,291 SNPs was used to genotype 453 *indica* accessions. The aims

Abbreviations: ABA, Abscisic acid; BLUP, Best linear unbiased prediction; cMLM, Compressed mixed linear model; CV, Coefficient of variation; DAP, Day after pollination; dCAPS, Derived cleaved amplified polymorphic sequences; EHH, Extended haplotype homozygosity; F_{ST} , Population differentiation; GA, Gibberellic acid; G \times E, Interactions of genotype and environment; GWAS, Genome-wide association study; H^2_B , Broad-sense heritability; HZ, Hangzhou; LD, Linkage disequilibrium; LS, Lingshui; MAE, Minor allele frequency; NJ, Neighbor-joining; PCA, Principal component analysis; PHS, Pre-harvest sprouting; PIC, Polymorphism information content; QTL, Quantitative trait locus; SE, Standard error; SNP, Single nucleotide polymorphism; SSR, Simple sequence repeat; XP-CLR, Cross-population composite likelihood ratio.

of our research were (1) to identify a substantial number of significantly associated SNPs and some putative genes potentially regulating seed dormancy in the whole panel; (2) to design dCAPS markers for different alleles of the associated SNPs for breeding applications; (3) to analyze the genetic diversity population structure and LD of the landraces and improved lines; and (4) to detect selective sweep regions using PIC ratio statistics between the landraces and improved lines.

MATERIALS AND METHODS

Plant Materials and SNP Genotypes

The association mapping population consisted of 453 *indica* accessions previously described in detail by Lu et al. (2015). According to the pedigree information, these accessions were classified into three types: landraces (266), improved lines (89) and foreign introduced lines (81) (Table S1). A rice landrace was defined as a “farm cultivar”, “traditional variety,” or “local variety” that had developed over a long time and adapted to the local natural and cultural environment. Rice landraces have always maintained rich genetic diversity. An improved line was defined as a “breeding variety” that was artificially selected by a breeder to meet a specific economic need, such as high yield, fine quality or multi-resistance. The genetic diversity of the improved lines was lower than that of the landraces. Therefore, the PIC ratios of the landraces (266) and improved lines (89) were used to detect whole-genome selective sweep signals. All accessions were planted in a randomized complete block design with three field replications and each line was planted in six rows with six hills per row, spacing at $\sim 20 \times 20$ cm, in Lingshui (LS; N 18°32', E 110°01') and Hangzhou (HZ; N 30°15', E 120°12') in 2014.

Rice genomic DNA was extracted from young leaf tissue and all accessions were genotyped using an Illumina custom-designed array containing 5,291 SNP markers following the Infinium HD Assay Ultra Protocol [https://support.illumina.com/downloads/infinium_hd_ultra_assay_protocol_guide_(11328087_b).html] (Illumina, Inc., San Diego, CA, USA) (Lu et al., 2015). Genotypes were called using the GenomeStudio software (Illumina, Inc.). The quality of each SNP was checked manually following a previous study (Yan et al., 2010). SNPs with tri- and tetra-allelic, low quality, high heterozygosity or a minor allele frequency (MAF) <5% were removed from the dataset. Finally, a total of 3,948 SNPs were chosen for further analysis (Table S2).

Seed Germination Evaluation and Data Analysis

The materials were cultivated under the same environmental conditions as much as possible. The heading date of each plant was recorded using the emergence of the first panicle from the flag leaf sheath as a standard. To reduce the marginal effect, four plants in the middle of each line were chosen as testers. The degree of seed dormancy was evaluated using the germination at seed maturity. To minimize the influence of conditions, ~ 100 filled seeds per line were harvested on the 30–40th day (in the case of the accessions) after heading, and then ~ 100 seeds were stored at 4°C to maintain seed freshness for germination

evaluation on the next day and another ~ 100 seeds per line were air-dried in greenhouse ($\sim 35^\circ\text{C}$) for 7 days and then treated at 45°C for 3 days to break dormancy. To test the germination, the seeds were wrapped in doubled sheets of 20×20 cm wetted absorbent filter paper, placed vertically, and germinated at 30°C and 100% relative humidity in the dark for 10 days. Germinated seeds were defined as those in which the length of shoot exceeded half the length of the seed. Germination (%) = number of germinated seeds \times 100/number of tested seeds. This experiment was repeated three times. The mean values of the three replications were used as the final phenotypic data. All seeds were intact grains.

Means, standard errors (SEs), broad-sense heritability (H_B^2), the percentage of phenotypic variation explained by population structure (R_Q^2) and interactions of genotype \times environment (G \times E) were calculated as in our previous report (Lu et al., 2015). The skewness, kurtosis and coefficient of variation (CV) were calculated using the functions Skew (), Kurt () and STDEV ()/AVERAGE () in Excel 2007, respectively.

Genome-Wide Association Analysis

To minimize the effects of environmental variation, best linear unbiased predictions (BLUPs) were performed using the R package lme4 (Bates et al., 2011) to estimate the phenotypic value for each line in the two environments (LS and HZ). The BLUP model can be described as

$$Y_{ijk} = L_k + E_i + R(E)_{ij} + (L \times E)_{ik} + \varepsilon_{ijk},$$

where Y_{ijk} is the observed phenotype for the k th line in the j th replicate of the i th environment; L_k is the random effect of the k th line; E_i is the random effect of the i th environment; $R(E)_{ij}$ is the random effect of the j th replicate in the i th environment; $(L \times E)_{ik}$ is the random interaction effect of the i th environment and the k th line, and ε_{ijk} is a random error following $N(0, \sigma_e^2)$.

GWASs were performed in TASSEL version 4.0 (Bradbury et al., 2007). The EMMA (Kang et al., 2008) and P3D (Zhang et al., 2010) algorithms were used to reduce computing time. The compressed mixed linear model (cMLM) with the population structure matrix (Q) and the relative kinship matrix (K) as covariates was used to reduce false-positive associations (Yu et al., 2006; Zhang et al., 2010). The Q matrix was calculated in our previous study (Lu et al., 2015) and the K matrix was generated using TASSEL version 4.0 (Bradbury et al., 2007) based on the 3,948 SNP markers. Four models including GLM, Q, K, and Q + K were used to evaluate type I errors (GLM model: no correction of population structure or familial relatedness; Q model: the population structure factor (Q matrix) was used as a covariate; K model: the familial relatedness effect (K matrix) was used as a covariate; Q + K model: both Q and K matrixes were used as covariates). The GLM and Q models were implemented using a general linear model program. However, the K and Q + K models were implemented in a mixed linear model program (Lu et al., 2015). The significance threshold for trait-marker associations was determined by Bonferroni correction ($\alpha = 1; 1/3,948 = 2.5\text{E-}04$) (Duggal et al., 2008; Yang et al., 2014). The Bonferroni-corrected threshold probability based on individual

tests is calculated to correct for multiple comparisons, using $1/N$ ($\alpha = 1$), where N is the number of individual trait-SNP combinations tested. Candidate genes were predicted within a 200-kb genomic region (± 100 kb of each significant SNP) from the Rice Haplotype Map Project Database (<http://202.127.18.221/RiceHap2/>) (Wang C. H. et al., 2014; Lu et al., 2015).

Candidate Gene Expression Profiles

The expression profiles of all candidate genes were analyzed according to the results of Davidson et al. (2012). Previous studies showed that seed dormancy-related genes had higher expression levels in the embryo, endosperm or seed (Sugimoto et al., 2010; Gu et al., 2011; Ye et al., 2015); thus, the expression data of Davidson et al. (2012) for these three tissues were used to screen the candidate genes. The screened candidate genes were then validated using the expression data in the Bio-Analytic Resource Plant Biology (BAR) database (<http://bar.utoronto.ca/>) and the Rice Genome Annotation Project database (<http://rice.plantbiology.msu.edu/>). SNPs located in the candidate genes and haplotype data were obtained from the RiceVarMap database (<http://ricevarmap.ncpgr.cn/>). Homologous gene identification was performed in the Rice Genome Annotation Project database, and then protein sequences were aligned using BLASTP in NCBI (<https://www.ncbi.nlm.nih.gov/>).

Linkage Disequilibrium Decay, Population Genetics and Polymorphism Information Content Analyses

To detect the regions of improvement sweeps, the 266 landraces and 89 improved lines were used to analyze LD decay, population structure and PIC ratios. The LD decay of the two panels was investigated using TASSEL version 4.0 (Bradbury et al., 2007). The LD decay rate was measured as the chromosomal distance at which the average pairwise correlation coefficient (r^2) dropped to half of its maximum value (Huang et al., 2010; Lu et al., 2015). A neighbor-joining (NJ) tree and principal component analysis (PCA) were used to infer the population structure. Based on Nei's genetic distance (Nei, 1972), the NJ tree was constructed using Powermarker 3.25 (Liu and Muse, 2005) and the PCA was performed using NTSYSpc version 2.1 (Rohlf, 2000). The F_{ST} among different subgroups and the PIC of each SNP marker were also evaluated using Powermarker 3.25 (Liu and Muse, 2005). The PIC ratio was used to test the improvement sweep regions according to differing genetic diversity of the two populations in the selected regions of the genome. The equation for the ratio can be expressed as follows: PIC ratio = $PIC_{(landrace)}/PIC_{(improvedline)}$. SNPs with PIC ratios > 3 were empirically considered improvement footprints.

RESULTS AND DISCUSSION

Diversity Panel and Phenotypic Variation

A rice diversity panel consisting of 453 *indica* accessions gathered from 20 rice-planting countries (Figure 1a) was used in our GWAS analyses. These accessions have rich seed dormancy variation, from deep to weak dormancy (Figures 1b,c). The germination of dormant seeds and dormancy-broken seeds were

used to confirm the existence of dormancy. The germination of dormancy-broken seeds was nearly 100%, which was significantly higher than that of dormant seeds regardless of environment (LS or HZ), indicating that all of the accessions had true dormancy characteristics and were viable (Figure 1d).

Germination of threshed seeds was used to test the degree of seed dormancy in our study. The $G \times E$ analysis indicated that environmental effects should not be ignored (Table 1). Thus, to minimize the effect of environment variation, BLUPs of the genetic effect for each line were used for the overall association analysis in the panel. The phenotypic variations of germination of threshed seeds in the two environments and predictions using the BLUP method are shown in Table 1 and Table S3. The mean germination were 34.7, 36.8, and 34.8% in LS, HZ and the BLUPs, respectively. The coefficients of variation were 87.8, 84.3, and 67.3%, respectively. Moreover, the ranges of germination were 0.0–98.2%, 0.0–97.4%, and 5.8–86.0%, respectively. These results suggested that the accessions had abundant phenotypic variation and were suitable for GWAS analysis (Pearson and Manolio, 2008). The positive skewness value ranged from 0.4 to 0.5, suggesting that the germination had a certain skewed distribution (Table 1, Figure 1b). The phenotypic variation explained by the population structure (R_Q^2) ranged from 45.7 to 61.2%; additionally, the broad-sense heritability (H_B^2) was 85.2%.

Rice has been found in archaeological sites dating to 8000 B.C. (Higham and Lu, 1998) and has been domesticated artificially and naturally from wild rice over a long period. Great changes have occurred in numerous traits, such as seed dormancy. Thus, one possible explanation for the skewed phenotype distribution was that the degree of seed dormancy in some cultivated rice accessions was significantly weakened to meet production needs during rice domestication (Sweeney and McCouch, 2007). The high R_Q^2 value suggested that the phenotypic variation was strongly affected by the population structure of the panel (Lu et al., 2015), and in further GWAS analyses, the population structure factor (Q matrix) should be taken into consideration to adjust the GWAS results. In addition, the relatively high H_B^2 indicated that genetic improvement of the trait was effective and could play a significant role in the breeding process in the future.

In recent years, GWASs in plants have become a new QTL detection strategy to unlock the genetic secrets of heritable traits through high-throughput genotyping technologies based on a large natural germplasm collection containing rich genetic and phenotypic variation (Huang et al., 2010). Compared with linkage mapping, GWASs use a germplasm mapping population and LD to improve the genetic mapping efficiency and mapping precision (Remington et al., 2001; Salvi and Tuberosa, 2005). Here, a large *indica* rice diversity panel containing a wealth of genetic diversity in seed dormancy (Figures 1b,c, Table 1) was used to perform a GWAS. Previous studies have also demonstrated that this *indica*-only population is well suited for association mapping (Lu et al., 2015, 2016).

Genome-Wide Analysis of Seed Dormancy

Our previous report showed that the *indica*-only panel could be classified into four populations (POPs) and a mixed subgroup (Mixed) (Figure S1), and relative kinship analysis

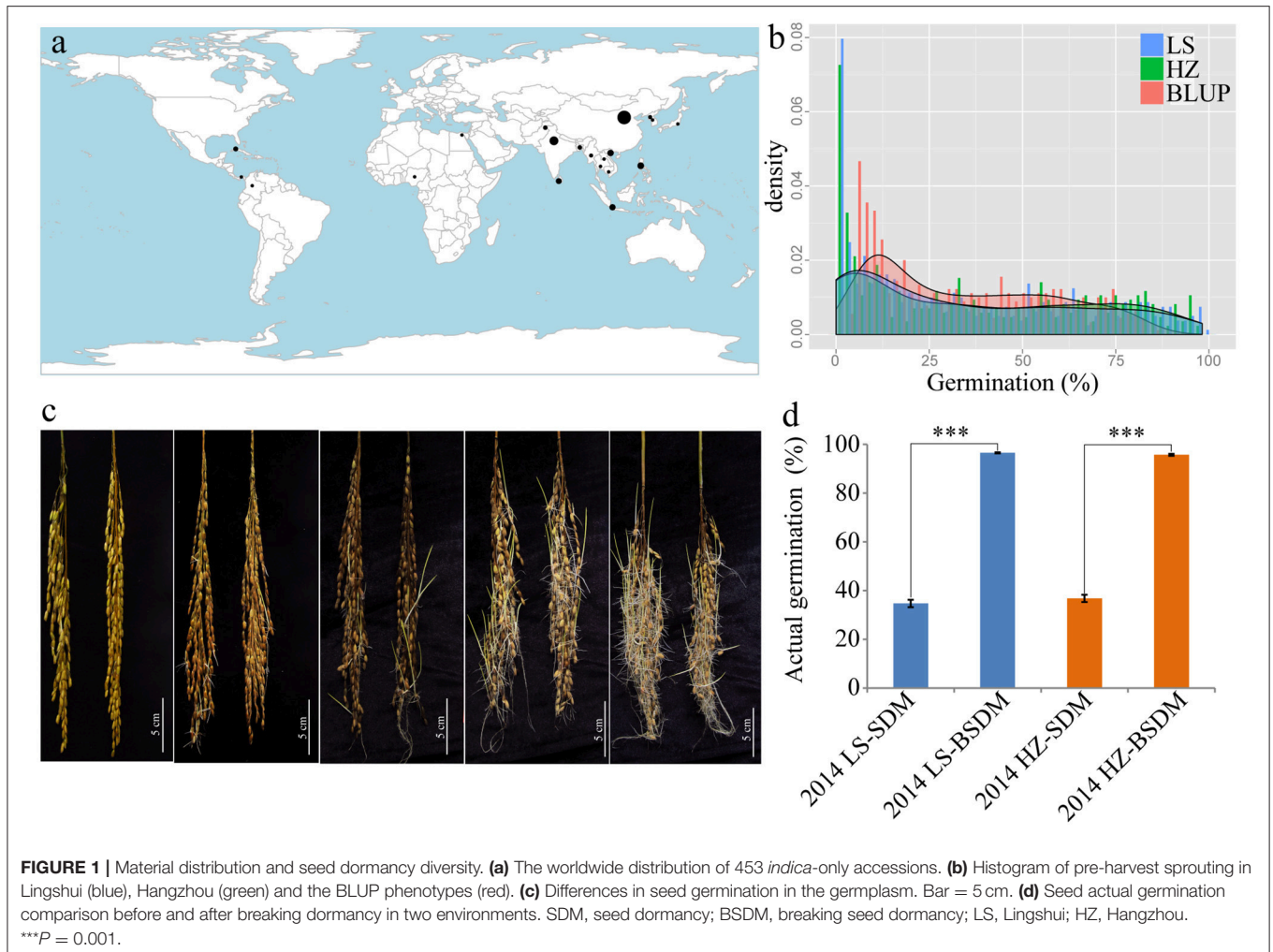


TABLE 1 | Phenotypic variation of pre-harvest seed germination in 453 *indica* accessions in two environments.

Trait	Germination (%)		
	2014-LS	2014-HZ	BLUP method
Env.			
Mean	34.7	36.8	34.8
SE	1.5	1.5	1.1
Min	0.0	0.0	5.8
Max	98.2	97.4	86.0
CV%	87.8	84.3	67.3
Skewness	0.5	0.4	0.4
Kurtosis	-1.1	-1.3	-1.1
R^2_Q %	45.7	61.2	60.0
H^2_B %		85.2	
G*E		***	

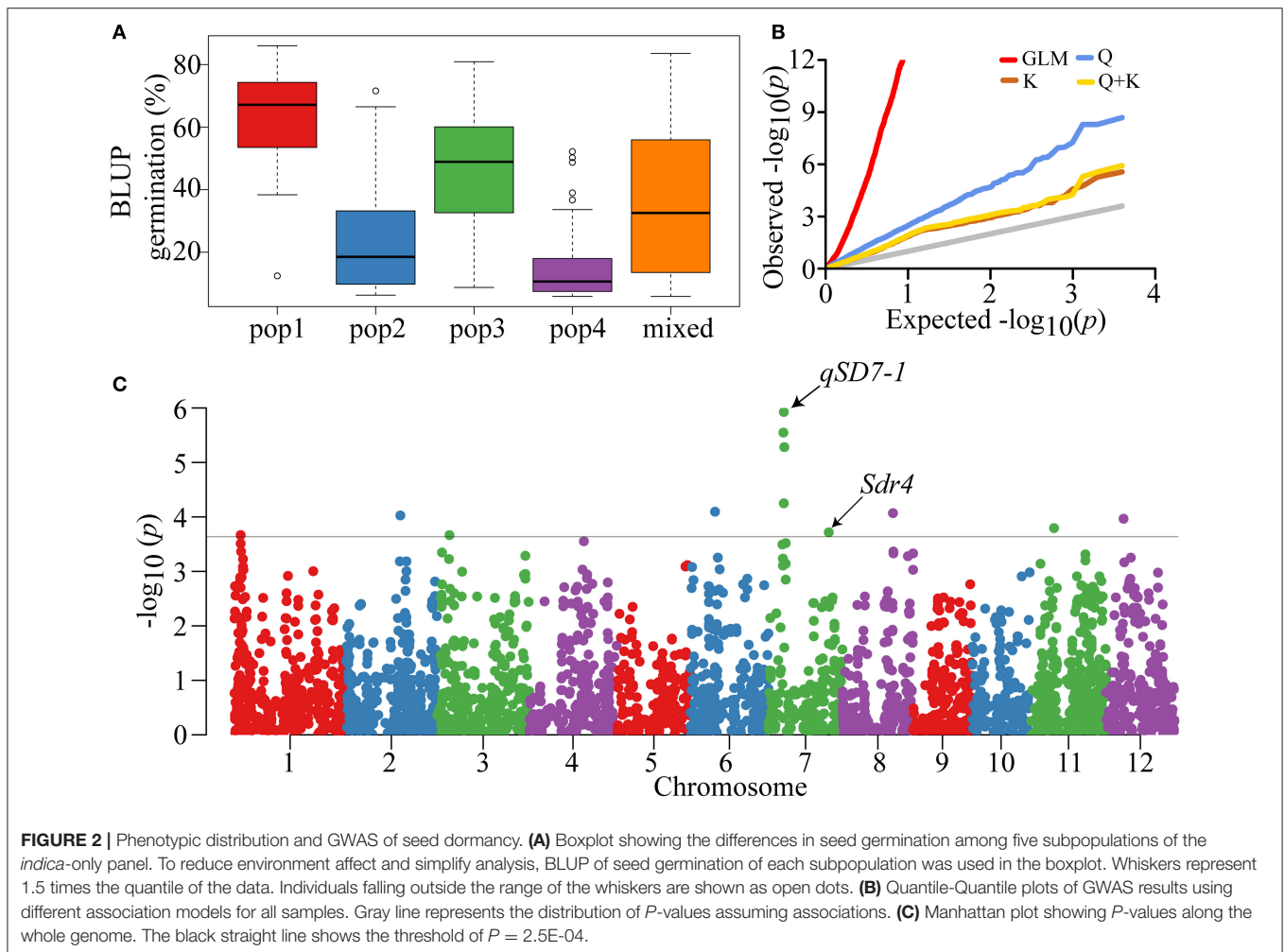
*** $P = 0.001$.

indicated that there was no or weak relatedness in the panel (Lu et al., 2015, 2016). Each accession was assigned to a corresponding group using Q component ≥ 0.6 as a threshold. The average germination was significantly different among the

five populations (Figure S1). In particular, POP1 had the highest value ($>60\%$) and POP4 had the lowest ($\sim 15\%$) (Figure 2A). This result was also supported by the high R^2_Q (Table 1). Taken together, these results highlighted the need to account for population structure and relative kinship when performing the subsequent GWAS analyses.

To evaluate the effects of the two elements and control false positive associations, four models (GLM, Q, K, and Q + K) were compared using a quantile-quantile (Q - Q) plot (Figure 2B). Compared with the GLM and Q models, the K and Q + K models showed great control of type I errors. Moreover, the Q + K model included both population structure and relatedness. Therefore, all further analyses were performed using the Q + K model with the cMLM.

The GWAS was conducted using the BLUPs of individual germination over the two environments. Through the GWAS, a total of 12 SNPs significantly associated with the seed dormancy trait were detected across eight chromosomes (Figure 2C). Four significant SNPs located close to each other were detected on the short arm of the chromosome 7 within a 260-kb region, among which the lead SNP (seq-rs3227, $P = 1.19E-06$) was used as the representative. As a result, nine peak



SNPs were retained (Table 2). The contribution of a single SNP to the phenotypic variation ranged from 3.5 to 5.8%, and together the SNPs explained 34.9% of the variation (Table 2). In addition, all SNPs except seq-rs5598 on chromosome 12 were adjacent to or overlapped with previously reported QTLs (Table 2). In particular, the strongest trait-associated SNP, seq-rs3227 ($P = 1.19E-06$), on chromosome 7 was located ~ 164.2 kb downstream of the cloned pleiotropic gene *qSD7-1* for seed dormancy, and another peak SNP, seq-rs3527 ($P = 1.93E-04$), resided ~ 3.7 kb downstream of *Sdr4*, which is involved in seed dormancy (Figure 2C; Figure S2) (Sugimoto et al., 2010; Gu et al., 2011).

In recent years, association studies have become a leading method to detect genes (or QTLs) underlying human diseases (Edwards et al., 2005) and agriculturally complex traits (Huang et al., 2010). However, the inflation of type I errors (false positive associations) caused by gene effects, allele frequencies, sample size and marker density is inevitable (Pe'er et al., 2006; Moonesinghe et al., 2007). Population structure and relative kinship are two common factors that determine the false positive rate (Zhang Z. et al., 2009). In our study, the high R^2_Q (Table 1), which was supported by the large germination

variation among the five different populations (Figure 2A), indicated that population stratification could cause some false positive associations when performing GWAS analyses. Thus, the cMLM (Q + K) (Zhang et al., 2016b) method was used to effectively eliminates false positive results by combining the two covariates simultaneously (Figure 2B).

In previous studies, many seed dormancy QTLs distributed on the 12 rice chromosomes have been identified using traditional linkage mapping (Gu et al., 2004, 2011; Guo et al., 2004; Wan et al., 2006; Sugimoto et al., 2010; Xie et al., 2011; Marzougui et al., 2012). In this study, the phenotypic variation explained by each individual significant SNP was $< 6\%$. This result demonstrated that rice seed dormancy is a typical quantitative trait with a minor genetic effect. Moreover, eight of the nine associated SNPs were detected previously in linkage analyses (Table 2) and two cloned seed dormancy genes, *qSD-7-1* and *Sdr4* on chromosome 7, were adjacent to the peak trait-associated SNPs seq-rs3227 and seq-rs3527, respectively (Table 2). Thus, the GWAS lead SNPs were largely confirmed by previous reports, showing that our results were reliable and implying that genome-wide association mapping is an effective strategy to uncover novel QTLs for complex agronomic traits based on a high density genetic map,

TABLE 2 | Summary of lead SNPs significantly associated with the seed dormancy trait.

Sig. SNPs	Allele	Chr.	Position (IRGSP v4)	P value	R ² (%)	Total R ² (%)	Known QTLs
seq-rs81	T/C	1	2,580,387	2.40E-04	3.5	34.9	RM6902 (Marzougui et al., 2012)
seq-rs1000	T/C	2	21,983,557	9.35E-05	3.9		RM525-RM318-RM240 (Xie et al., 2011)
seq-rs1445	G/A	3	5,178,318	2.36E-04	3.5		RM231 (Wan et al., 2006)
seq-rs2896	C/A	6	10,198,431	7.96E-05	4.0		RM5963 (Marzougui et al., 2012); RM549 (Gu et al., 2004)
seq-rs3227	G/A	7	6,265,668	1.19E-06	5.8		<i>qSD7-1</i> (Gu et al., 2011); RM180 (Gu et al., 2004)
seq-rs3527	C/A	7	24,461,489	1.93E-04	3.6		FHS7.0 (Magwa et al., 2016); <i>Sdr4</i> (Sugimoto et al., 2010); RM234 (Wan et al., 2006); RM346 (Gu et al., 2004)
seq-rs3939	A/G	8	20,453,214	8.44E-05	4.0		FHS8.1 (Magwa et al., 2016); RM531 (Gu et al., 2004)
seq-rs4917	A/G	11	9,746,917	1.60E-04	4.1		<i>qSD-11</i> (Guo et al., 2004)
seq-rs5598	A/C	12	7,132,922	1.07E-04	3.9		

although only one SNP, seq-rs5598 ($P = 1.07E-04$), was not detected in previous reports.

Seed dormancy is easily affected by multiple environmental factors, and seed maturing time is one of the most important influencing factors. However, it is difficult to investigate because there is no effective standard to define or evaluate when the seeds are mature. Generally speaking, seed that turns yellow and hardened means maturity. But this is a continuous process of changes and it is hard to record a specific time. In this study, we harvested filled seed on the 30–40th day after heading in the case of each accession by empiricism to reduce the effects of seed maturing time. In conclusion, the relationship between seed dormancy and seed maturing time should be explored in depth. In addition, how to define and determine seed maturing time that may be related to grain-filling rate will also be a valuable topic to deep study in future.

Trait-Associated SNPs Effects and Molecular Breeding Application

The phenotypic differences between the two alleles of each of the strongest trait-associated SNPs are summarized in Table S4. Using the absolute value as the standard, the deviation values ranged from ~0.67 to ~25.4% for each SNP. Moreover, the deviation values of five SNPs reached significant ($P \leq 0.05$) or highly significant ($P \leq 0.01$ or 0.001) levels between the two alleles of each SNP marker (Figure 3). Among them, the absolute values of four SNPs were close to or more than 10%, especially for the seq-rs3527 ($P = 1.93E-04$), which was ~25.4% (Figure 3B; Table S4). The results suggested that all these trait-associated SNPs could be used efficiently for molecular breeding in the future.

To confirm this, the elite alleles with positive effects were used to test the effectiveness of pyramid breeding. Without considering the effects of interactions among these lead SNPs and environmental influences, the more elite alleles that were pyramided in a variety, the higher the germination increase (Figure 3E). This result indicated that pyramiding of favorable alleles could attenuate seed dormancy. In practice, according to their breeding goals, breeders could introduce different numbers of favorable alleles into different varieties to modify the seed dormancy trait. In the GWAS panel, most of the accessions

carried four to six favorable alleles and had moderate germination ranging from 30.8 to 42.4%. Only a few of accessions had an extreme germination; <18.1% or larger than 48.3% (Figure 3E). This result suggested that a suitable number of favorable alleles have been maintained in these accessions through artificial phenotypic selection in the process of rice breeding to meet cultivation needs. Accessions with extreme phenotypes (very high or low PHS) do not suit human needs and are gradually eliminated during the breeding process, but these accessions may be good materials for genetic research. Since LS and HZ are of different latitudes, the day-length and temperature are quite different. Which environment would be more suitable for the favorable allele mining for the *indica* breeding practice? Alleles favorable under both locations were stably expressed would be widely adopted by more breeders throughout the country. However, those mined specifically at one location would also offer useful information for the breeding work under similar conditions. For example, the changes of day-length and temperature are more obvious in HZ other than that in LS throughout the year. Consequently, the phenotypic effects of some favorable alleles would be more apparent and easier to observe, especially for those that sensitive to the day-length and temperature.

How can these useful lead SNPs be applied to molecular breeding? Although the flanking sequence of each SNP (Table S2) can be used to detect the SNPs for breeding with next-generation sequencing technology, to some extent, this method is inconvenient. Thus, a dCAPS marker for each SNP was designed using dCAPS Finder 2.0 (Neff et al., 2002) (Table S5). These dCAPS markers can clearly distinguish the genotypes of the corresponding SNPs in polypropylene gel electrophoresis (Figure S3). Generally speaking, these markers will be beneficial for molecular marker-assisted selection breeding in the future.

Candidate Gene Prediction and Expression Profiling

The flanking regions within a 200-kb window (± 100 kb) of the lead SNPs were searched to identify candidate genes in the Rice Haplotype Map Project Database (Wang C. et al., 2014; Lu et al., 2016). For the two SNPs on chromosome 7, seq-rs3227 ($P = 1.19E-06$) and seq-rs3527 ($P = 1.93E-04$), two

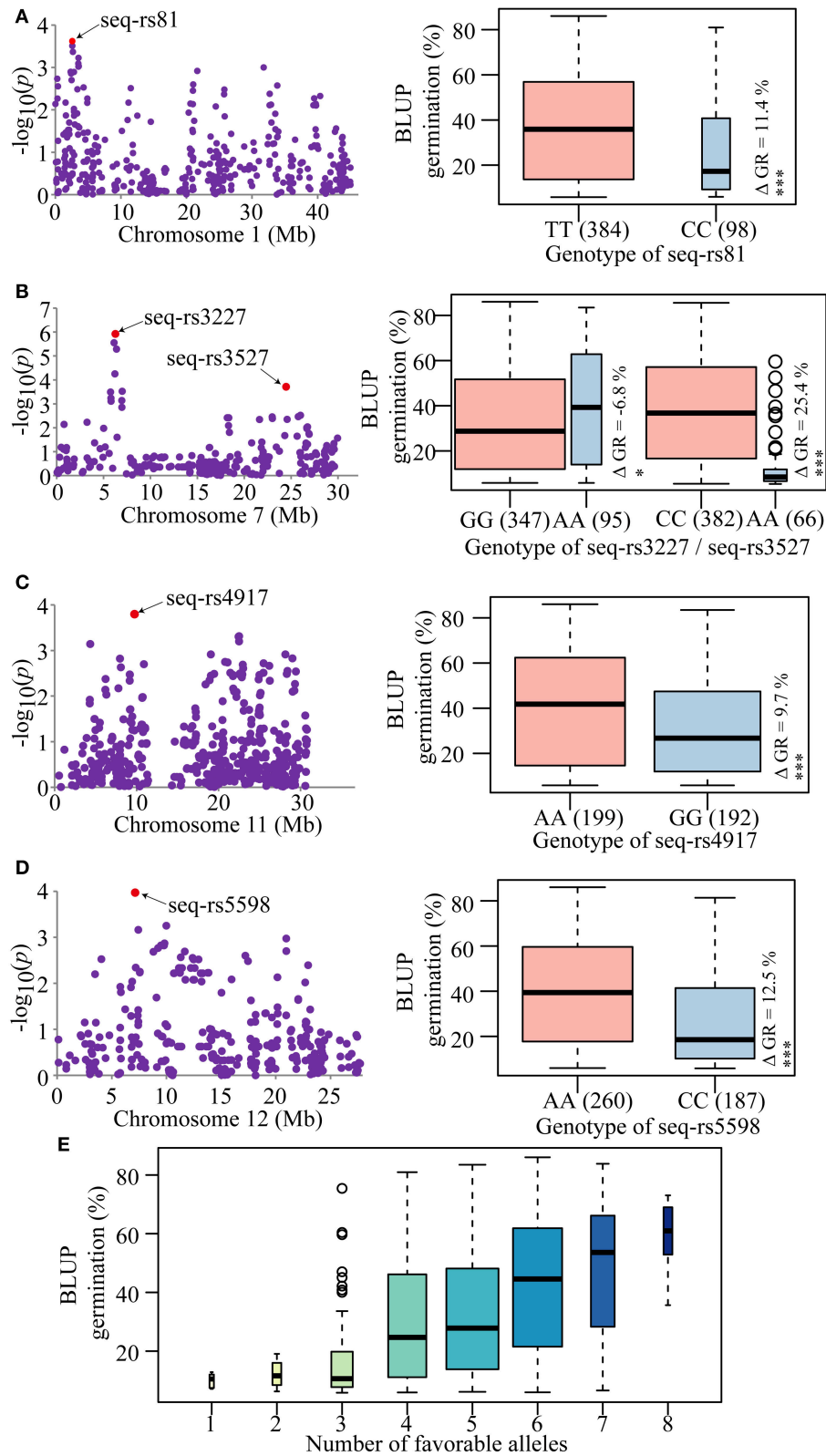


FIGURE 3 | Trait-associated SNPs and pyramiding of favorable alleles. **(A–D)** Candidate SNPs associated with seed dormancy and phenotypic differences between the two alleles of each SNP. Red dots represent the lead SNPs of candidate regions. The boxplots on the right show the distribution of average germination estimated (Continued)

FIGURE 3 | by BLUP for each SNP allele. The number of individuals for each allele is given in parenthesis and is represented by the width of the box. Whiskers represent 1.5 times the quantile of the data. Individuals falling outside the range of the whiskers are shown as open dots. Differences in means are shown by Δ GR. * $P = 0.05$ and *** $P = 0.001$, respectively. **(E)** Pyramid effects for different numbers of favorable alleles of the candidate SNPs. The X-axis represents the number of elite alleles carried by the accessions and the Y-axis represents the trait mean value estimated by BLUP. The width of each boxplot represents the number of accessions.

known seed dormancy-related genes, *qSD7-1* (Gu et al., 2011) and *Sdr4* (Sugimoto et al., 2010), respectively, were located in the corresponding region (Table 2, Figure S2). For the other seven SNPs, a total of 212 candidate genes were identified, which are summarized in Table S6.

To identify the most promising candidate genes, the expression levels in embryo, endosperm and seed tissues, which were downloaded from the results of Davidson et al. (2012), were used as a screening reference. Interestingly, *qSD7-1* had a high expression level in the seed at 5 days after pollination (DAP). By comparison, the *Sdr4* had high expression levels in the embryo at 25 DAP, endosperm at 25 DAP, and seeds at 5 and 10 DAP (Figures 4A,B, Table S6). These results were consistent with previous reports (Sugimoto et al., 2010; Gu et al., 2011; Ye et al., 2015). After searching the expression data for the 212 candidate genes, the expression patterns of eight genes were found to be similar to those of *qSD7-1* or *Sdr4* (Figures 4A,B, Table S6). Then, expression data from the BAR database and the rice genome annotation project database were used to verify the eight genes (Figure 4C, Table S6). The candidate gene LOC_Os03g10110, which encoded a cupin domain-containing protein, had high expression levels in seed tissues at different stages (Figures 4B,C, Table S6). Moreover, the expression pattern of this gene was highly consistent with that of *Sdr4*, which has very high expression levels in seeds at 5 and 10 DAP, embryos at 25 DAP, and endosperm at 25 DAP (Figure 4D, Table S6). Homology analysis indicated that the gene was homolog to grmzm2g078441, a gene in the cupin family of unknown function that is highly expressed in embryo tissues in maize (Teoh et al., 2013). A previous report suggested that cupins comprise a superfamily of functionally diverse proteins that include germins and plant storage proteins (Dunwell, 1998). Storage proteins may be important in plants during seed germination and seedling growth (Shewry et al., 1995). The RiceVarMap database (<http://ricevarmap.ncpgr.cn/>) showed that a total of 34 SNPs were located within this gene (Table S7) and that 15 SNPs were non-synonymous mutations (Figure 4E). Haplotype network analysis indicated that most *indica* accessions could be classified into haplotype groups I and III, while the majority of *japonica* accessions were assigned to haplotype group II (Figure 4F, Table S8). This result suggested that the haplotype of the candidate gene displayed some *indica-japonica* specificity.

Population Division, Nucleotide Diversity and Linkage Disequilibrium between Improved Lines and Landraces

Rice improvement is the outcome of continuous artificial selection to enhance the adaptation of the plant to fit human needs. Consequently, various rice traits, such as plant type and

seed shape, have been changed dramatically (Figures S4A,B). The successive selection effect also alters the genetic diversity at the genomic level, resulting in a distorted pattern of genetic variation and LD decay.

To identify selective sweeps across the whole rice genome, the 89 improved lines and 266 landraces, most of which were widely distributed in southern China, were used to detect selective sweep signals (Figures S4C,D). To better understand the population stratification and geographic structure diversity, a NJ tree was constructed and PCA was performed to illustrate the relatedness among the accessions. The results indicated that the improved lines comprised four subgroups, among which PC1 and PC2 accounted for 22.0 and 8.4% of the genetic variation, respectively (Figures 5A,B). However, the landraces were classified into five subgroups, among which PC1 and PC2 explained 18.2 and 10.0% of the genetic variation, respectively (Figures 5C,D). The F_{ST} ranged from 0.07 to 0.47 among all nine subgroups, indicating strong population differentiation among some of the subgroups (Figure S5). The F_{ST} averaged 0.17 among the subgroups of the improved lines, suggesting a moderate level of differentiation, and was estimated at 0.26 on average among the landrace subgroups, implying greater population differentiation than in the improved lines (Figure S5). These averages were close to previously published values in rice (Wang C. H. et al., 2014; Lu et al., 2015), but much less than that between *indica* and *japonica* ($F_{ST} = 0.55$) (Huang et al., 2010), and a little bit larger than in soybean (Wen et al., 2014), maize (Yang et al., 2011), and sesame (Wei et al., 2015).

To further evaluate the degree of human selection, the gene diversity, PIC and LD decay were quantified for the two populations. Gene diversity and PIC values for the landraces were both significantly higher than in the improved lines (Figures 6A,B). These estimations were close to previous estimates from 926 SNPs in maize (Yang et al., 2011), but much smaller than those calculated from simple sequence repeat (SSR) markers in rice (Wang C. H. et al., 2014) and spring barley landraces (Pasam et al., 2014). This result demonstrated that the landraces had retained more genetic diversity than the improved lines. Because increased LD is another hallmark of artificial selection in rice, the LD decay rates between the two populations were compared. The extent of LD decay increased from 163.3 kb for the landraces to 352.4 kb for the improved lines (Figure 6C). The larger LD decay distance in the improved lines may have been caused by a loss of genetic diversity and a low frequency of genetic recombination because of human selection forces during improvement (Lam et al., 2010). The extent of LD for the landraces was similar to that in a previous evaluation in rice (Huang et al., 2010; Lu et al., 2015), but much greater than that in maize (Remington et al., 2001).

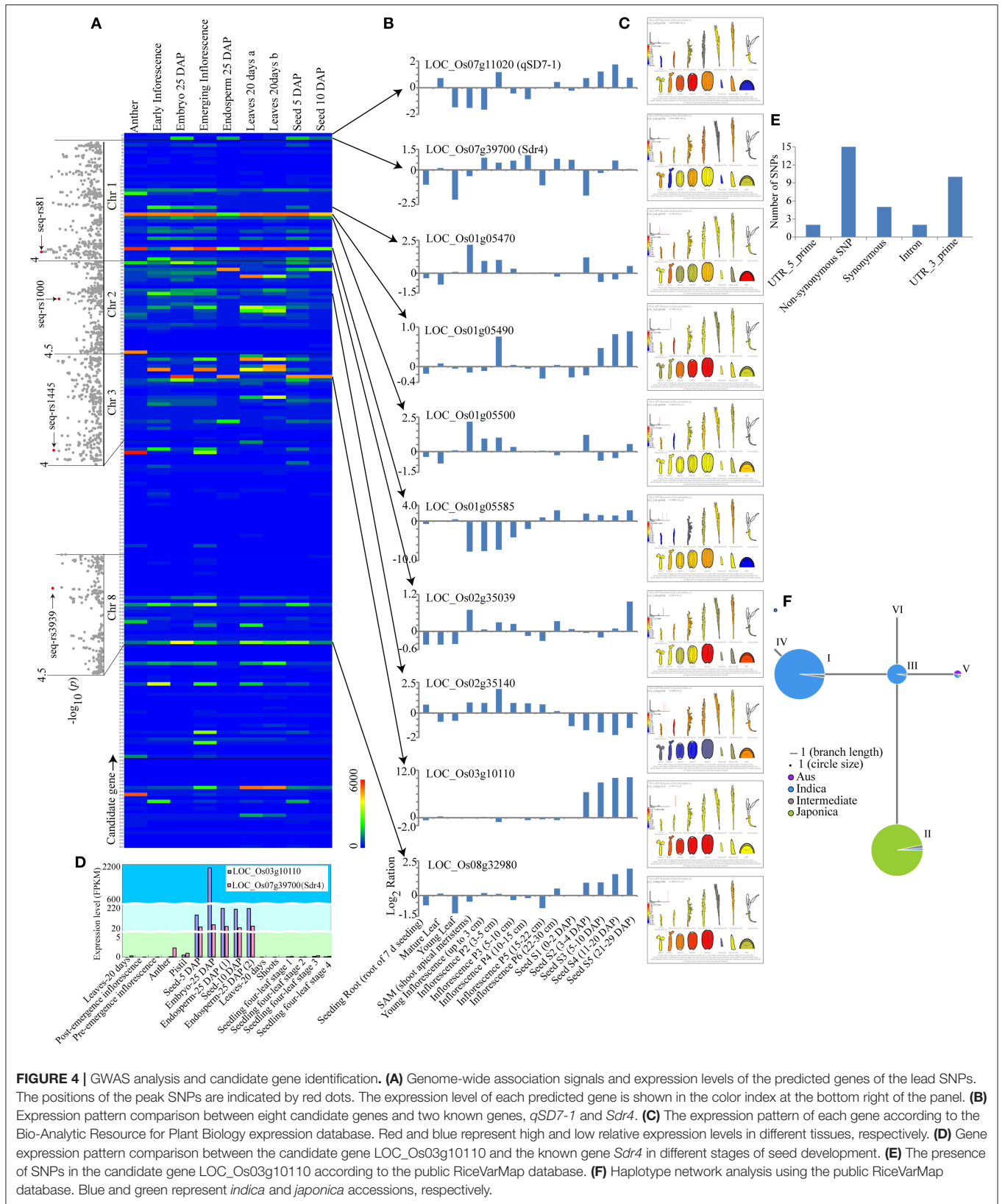
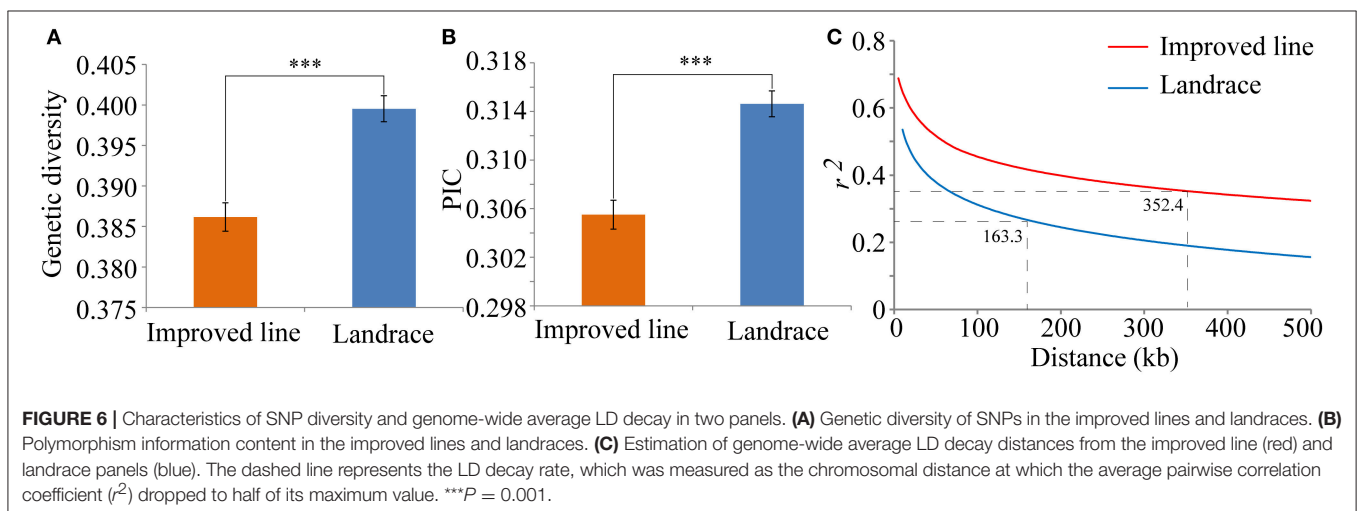
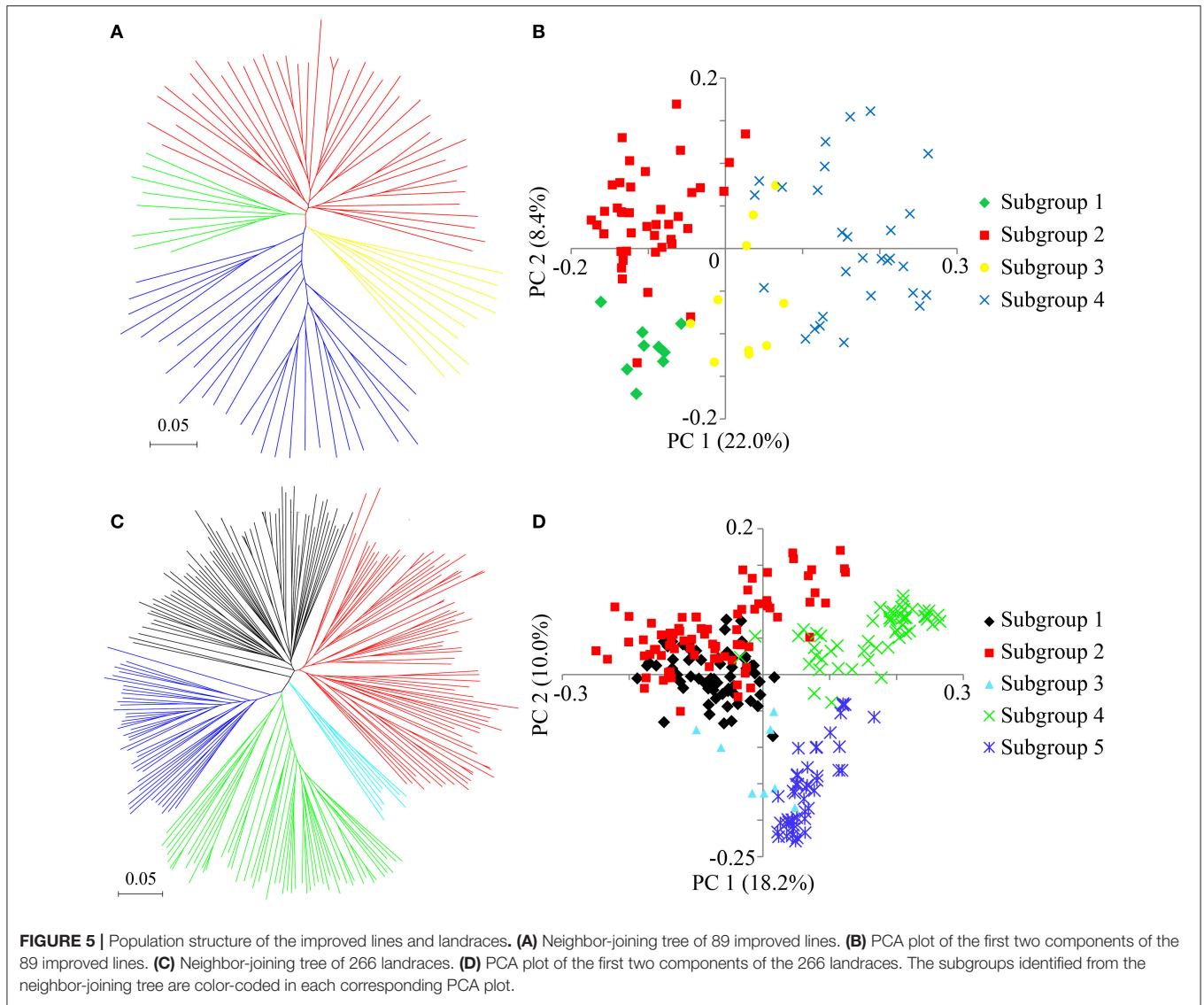
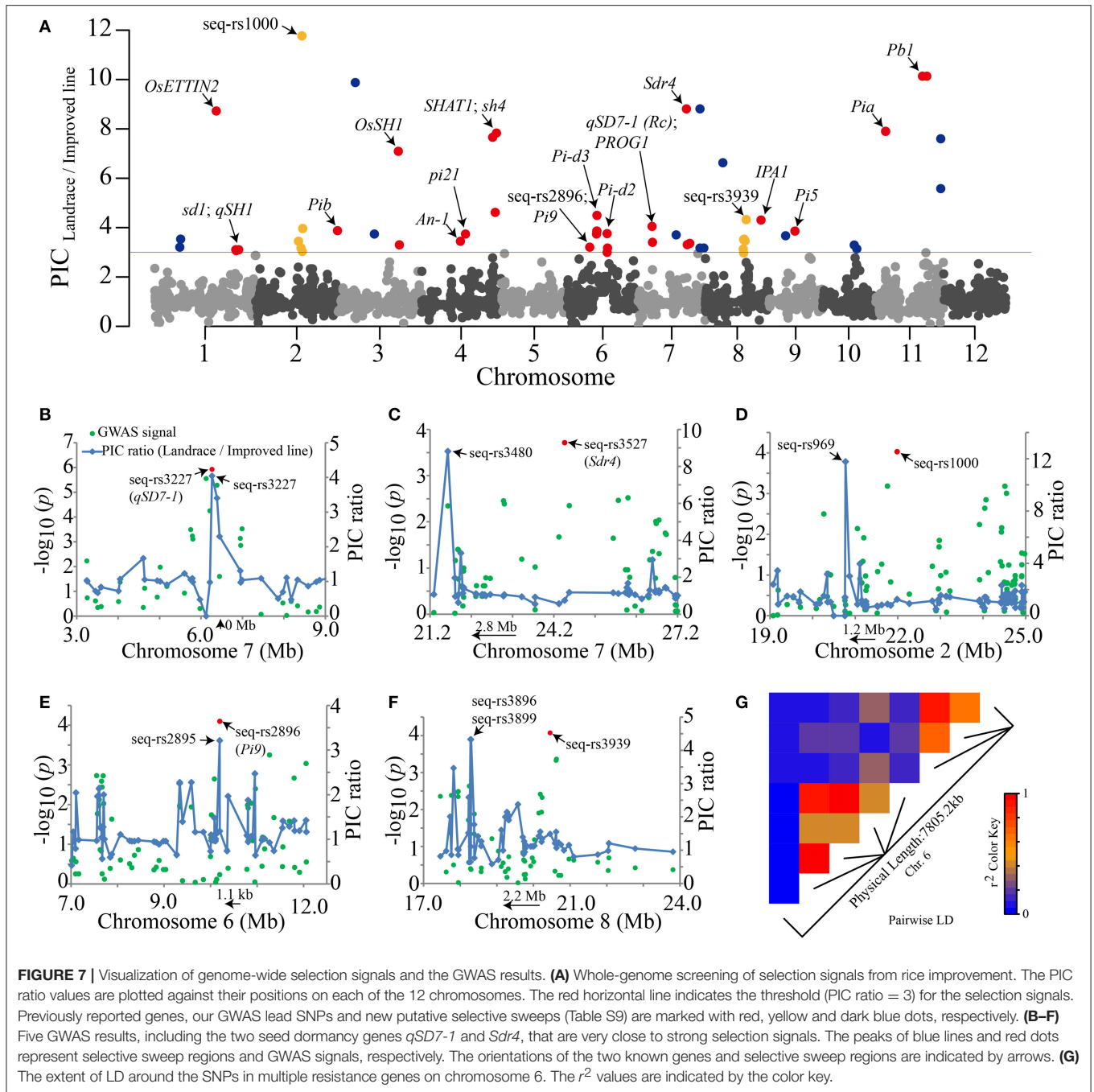


FIGURE 4 | GWAS analysis and candidate gene identification. **(A)** Genome-wide association signals and expression levels of the predicted genes of the lead SNPs. The positions of the peak SNPs are indicated by red dots. The expression level of each predicted gene is shown in the color index at the bottom right of the panel. **(B)** Expression pattern comparison between eight candidate genes and two known genes, *qSD7-1* and *Sdr4*. **(C)** The expression pattern of each gene according to the Bio-Analytic Resource for Plant Biology expression database. Red and blue represent high and low relative expression levels in different tissues, respectively. **(D)** Gene expression pattern comparison between the candidate gene *LOC_Os03g10110* and the known gene *Sdr4* in different stages of seed development. **(E)** The presence of SNPs in the candidate gene *LOC_Os03g10110* according to the public RiceVarMap database. **(F)** Haplotype network analysis using the public RiceVarMap database. Blue and green represent *indica* and *japonica* accessions, respectively.





Genome Wide Selective Sweep Signal Scan

Artificial selection has probably changed the nucleotide diversity within the genomes of cultivars. To detect the genomic regions most affected by artificial selection during rice improvement, PIC ratios between the landraces and improved lines were screened throughout the whole genome (Figure 7A). PIC ratios >3.0 (the top ~1% of all values) were retained as an empirical threshold. After screening, the PIC ratios of 57 SNPs exceeded the cutoff. Peak signals within a $\sim\pm 1.5$ -Mb window were grouped into a single DNA region because the selection effect leads to greater LD

decay and extended haplotype structure (Qanbari et al., 2014). Finally, 27 selection DNA regions were identified (Table S9). Among these regions, 18 known domestication-related genes resided in 15 corresponding selection regions, such as the “green revolution” gene *sd1* (Sasaki et al., 2002), the seed shattering-related gene *sh4* (Li et al., 2006), *qSD7-1* and *Sdr4* for seed dormancy (Sugimoto et al., 2010; Gu et al., 2011) and *IPA1* for ideal plant architecture (Jiao et al., 2010) (Figure 7A, Table S9).

Interestingly, in our GWAS, the *qSD7-1* gene for seed dormancy was located near the most significant SNP seq-rs3227

($P = 1.19E-06$) (Table 2). This SNP also showed strong selection signals with a highest PIC ratio value of ~ 4.0 within a 6-Mb genome region (Figure 7B). Notably, another selection signal, seq-rs3480 (PIC ratio = 8.8), appeared near the seed dormancy-related gene *Sdr4*, which resided upstream of the GWAS signal of seq-rs3527 ($P = 1.93E-04$) (Table 2) within a ~ 2.8 Mb physical distance (Figure 7C). More importantly, there were three strong selection signals in the vicinity of the corresponding three peak SNP previously identified by the GWAS method (Table 2, Figures 7D–F). In particular, one of the GWAS peak SNPs (seq-rs2896, $P = 7.96E-05$) was located just 1.1 kb downstream of the strongest selection SNPs (seq-rs2895, PIC ratio = 3.2) (Figure 7E). In addition, on chromosome 6, there were eight selection signals adjacent to rice blast resistance genes (Figure 7A, Table S9). The pairwise LD values showed that most of these were not located in one haplotype block ($r^2 < 0.8$) (Figure 7G), implying that rice resistance-related traits may also have experienced strong artificial selection during rice improvement (Figure 7E).

Taken together, our results suggested most of the selective sweep regions ($\sim 66.7\%$) were adjacent to both known domestication-related genes and SNPs we identified previously by GWAS (Table 2, Figure 7, Table S9). These results indicated that genome-wide screening to detect selective sweeps using PIC ratios between two populations with different degrees of human selection can be used to identify DNA regions potentially related to artificial selection in rice. Additionally, the accuracy of our previous GWAS results (Table 2) was further validated by these selective sweep regions. However, no selection signals were detected on chromosomes 5 and 12. There are two possible reasons for this. On the one hand, a lack of sufficient polymorphic SNP markers in these specific genome regions among the two panels may have led to poor detection capability. On the other hand, weak selective pressure on these two chromosomes may have limited our power to detect selection signals.

CONCLUSIONS

In this study, a total of nine known and new SNPs associated with rice seed dormancy were identified via GWAS. dCAPS markers were designed to accelerate the molecular breeding of rice dormancy. Moreover, 212 candidate genes were identified. The expression profiles and haplotype network data from public databases revealed eight genes, especially LOC_Os03g10110, which has a maize homolog involved in embryo development, as candidate regulators for further investigations to verify their biological functions. A genome-wide screen to detect artificial

REFERENCES

- Alonso-Blanco, C., Bentsink, L., Hanhart, C. J., Blankestijn-de Vries, H., and Koornneef, M. (2003). Analysis of natural allelic variation at seed dormancy loci of *Arabidopsis thaliana*. *Genetics* 164, 711–729. Available online at: <http://www.genetics.org/content/164/2/711>
- Amiguet-Vercher, A., Santuari, L., Gonzalez-Guzman, M., Depuydt, S., Rodriguez, P. L., and Hardtke, C. S. (2015). The *IBO* germination quantitative trait locus

selection signals identified 27 selection DNA regions. Among them, 15 were adjacent to known domestication-related genes and three strong selection signals were located near GWAS lead SNPs. These results not only further verify the accuracy of our GWAS findings but also suggest that genome-wide screening for selective sweeps can be used to identify new improvement-related DNA regions, although the phenotypes are unknown. This study enhances our knowledge of the genetic variation in rice seed dormancy, and the new GWAS SNPs will provide real benefits for genomic selection in breeding programs. More importantly, the genomic consequences of improvement footprints will enable the detection of domestication-related traits.

AVAILABILITY OF DATA AND MATERIAL

The SNP dataset used during this study is available in the in the Dryad digital repository (<http://dx.doi.org/10.5061/dryad.cp25h>). Any other datasets used and/or analyzed during the current study available from the corresponding author on reasonable request.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: QL, XN, and XW. Performed the experiments: QL, XN, MZ, CW, QX, and YF. Analyzed the data: QL, YY, SW, XY, HY, and YW. Wrote the paper: QL. Revised the manuscript: XC, XL, and XW. All authors read and approved the final manuscript.

FUNDING

This research was supported by grants from the Ministry of Science and Technology of China (2016YFD0100101-02) and the Science and Technology Planning Project of Guangdong Province (2015B020231006, 2016B020201003).

ACKNOWLEDGMENTS

We sincerely thank to Dr. Xuehui Huang and Dr. Bin Han (National Center for Gene Research, Chinese Academy of Sciences) for their excellent discussions and constructive suggestion.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2017.02213/full#supplementary-material>

encodes a phosphatase 2C-related variant with a non-synonymous amino acid change that interferes with abscisic acid signaling. *New Phytol.* 205, 1076–1082. doi: 10.1111/nph.13225

- Asano, K., Yamasaki, M., Takuno, S., Miura, K., Katagiri, S., Ito, T., et al. (2011). Artificial selection for a green revolution gene during *japonica* rice domestication. *Proc. Natl. Acad. Sci. U.S.A.* 108, 1034–1039. doi: 10.1073/pnas.1019490108

- Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., et al. (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631. doi: 10.1038/nature08800
- Basbous-Serhal, I., Leymarie, J., and Bailly, C. (2016). Fluctuation of *Arabidopsis* seed dormancy with relative humidity and temperature during dry storage. *J. Exp. Bot.* 67, 119–130. doi: 10.1093/jxb/erv439
- Bates, D., Mächler, M., and Dai, B. (2011). *lme4: Linear Mixed-Effects Models using Eigen and SVD*. R Package Version 0.999375-28. Available online at: <http://lme4.r-forge.r-project.org>
- Begum, H., Spindel, J. E., Lalusin, A., Borromeo, T., Gregorio, G., Hernandez, J., et al. (2015). Genome-wide association mapping for yield and other agronomic traits in an elite breeding population of tropical rice (*Oryza sativa*). *PLoS ONE* 10:e0119873. doi: 10.1371/journal.pone.0119873
- Bewley, J. D. (1997). Seed Germination and dormancy. *Plant Cell* 9, 1055–1066. doi: 10.1105/tpc.9.7.1055
- Bewley, J. D., and Black, M. (1982). *Physiology and Biochemistry of Seeds in Relation to Germination*. Chapter 2: 61–81. Berlin; Heidelberg; New York, NY: Springer-Verlag.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23, 2633–2635. doi: 10.1093/bioinformatics/btm308
- Cai, H. W., and Morishima, H. (2000). Genomic regions affecting seed shattering and seed dormancy in rice. *Theor. Appl. Genet.* 100, 840–846. doi: 10.1007/s001220051360
- Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. doi: 10.1101/gr.100545.109
- Davidson, R. M., Gowda, M., Moghe, G., Lin, H., Vaillancourt, B., Shiu, S. H., et al. (2012). Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution. *Plant J.* 71, 492–502. doi: 10.1111/j.1365-313X.2012.05005.x
- Duggal, P., Gillanders, E. M., Holmes, T. N., and Bailey-Wilson, J. E. (2008). Establishing an adjusted *p*-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics* 9:516. doi: 10.1186/1471-2164-9-516
- Duijvesteijn, N., Knol, E. F., Merks, J. W., Crooijmans, R. P., Groenen, M. A., Bovenhuis, H., et al. (2010). A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. *BMC Genet.* 11:42. doi: 10.1186/1471-2156-11-42
- Dunwell, J. M. (1998). Cupins: a new superfamily of functionally diverse proteins that include germins and plant storage proteins. *Biotechnol. Genet. Eng. Rev.* 15, 1–32. doi: 10.1080/02648725.1998.10647950
- Edwards, A. O., Ritter, R. III., Abel, K. J., Manning, A., Panhuysen, C., and Farrer, L. A. (2005). Complement factor H polymorphism and age-related macular degeneration. *Science* 308, 421–423. doi: 10.1126/science.1110189
- Fang, J., Chai, C., Qian, Q., Li, C., Tang, J., Sun, L., et al. (2008). Mutations of genes in synthesis of the carotenoid precursors of ABA lead to pre-harvest sprouting and photo-oxidation in rice. *Plant J.* 54, 177–189. doi: 10.1111/j.1365-313X.2008.03411.x
- Feng, F., Liu, P., Hong, D., and Yang, G. (2009). A major QTL associated with preharvest sprouting in rapeseed (*Brassica napus* L.). *Euphytica* 169, 57–68. doi: 10.1007/s10681-009-9921-8
- Fennimore, S. A., Nyquist, W. E., Shaner, G. E., Doerge, R. W., and Foley, M. E. (1999). A genetic model and molecular markers for wild oat (*Avena fatua* L.) seed dormancy. *Theor. Appl. Genet.* 99, 711–718. doi: 10.1007/s001220051288
- Flintham, J., Adlam, R., Bassoi, M., Holdsworth, M., and Gale, M. (2002). Mapping genes for resistance to sprouting damage in wheat. *Euphytica* 126, 39–45. doi: 10.1023/A:1019632008244
- Gandhi, S. D., Heesacker, A. F., Freeman, C. A., Argyris, J., Bradford, K., and Knapp, S. J. (2005). The self-incompatibility locus (S) and quantitative trait loci for self-pollination and seed dormancy in sunflower. *Theor. Appl. Genet.* 111, 619–629. doi: 10.1007/s00122-005-1934-7
- Gao, W., Clancy, J. A., Han, F., Prada, D., Kleinhofs, A., and Ullrich, S. E. (2003). Molecular dissection of a dormancy QTL region near the chromosome 7 (5H) L telomere in barley. *Theor. Appl. Genet.* 107, 552–559. doi: 10.1007/s00122-003-1281-5
- Gawenda, I., Thorwarth, P., Günther, T., Ordon, F., and Schmid, K. J. (2015). Genome-wide association studies in elite varieties of German winter barley using single-marker and haplotype-based methods. *Plant Breed.* 134, 28–39. doi: 10.1111/pbr.12237
- Gu, X. Y., Foley, M. E., Horvath, D. P., Anderson, J. V., Feng, J., Zhang, L., et al. (2011). Association between seed dormancy and pericarp color is controlled by a pleiotropic gene that regulates abscisic acid and flavonoid synthesis in weedy red rice. *Genetics* 189, 1515–1524. doi: 10.1534/genetics.111.131169
- Gu, X. Y., Kianian, S. F., and Foley, M. E. (2004). Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*Oryza sativa*). *Genetics* 166, 1503–1516. doi: 10.1534/genetics.166.3.1503
- Gu, X. Y., Liu, T., Feng, J., Suttle, J. C., and Gibbons, J. (2010). The *qSD12* underlying gene promotes abscisic acid accumulation in early developing seeds to induce primary dormancy in rice. *Plant Mol. Biol.* 73, 97–104. doi: 10.1007/s11103-009-9555-1
- Gu, X. Y., Zhang, J., Heng, Y., Zhang, L., and Feng, J. (2015). Genotyping of endosperms to determine seed dormancy genes regulating germination through embryonic, endospermic, or maternal tissues in rice. *G3 (Bethesda)* 5, 183–193. doi: 10.1534/g3.114.015362
- Guo, L., Zhu, L., Xu, Y., Zeng, D., Wu, P., and Qian, Q. (2004). QTL analysis of seed dormancy in rice (*Oryza sativa* L.). *Euphytica* 140, 155–162. doi: 10.1007/s10681-004-2293-1
- Higham, C., and Lu, T. L. D. (1998). The origins and dispersal of rice cultivation. *Antiquity* 72, 867–877. doi: 10.1017/S0003598X00087500
- Hu, W. M., Ma, H. S., Fan, L. J., and Ruan, S. L. (2003). Characteristics of pre-harvest sprouting in sterile lines in hybrid rice seeds production. *Acta Agron. Sin.* 29, 441–446 (in Chinese with English abstract). Available online at: <http://zwxb.chinacrops.org/EN/abstract/abstract1863.shtml>
- Huang, X., Kurata, N., Wei, X., Wang, Z. X., Wang, A., Zhao, Q., et al. (2012). A map of rice genome variation reveals the origin of cultivated rice. *Nature* 490, 497–501. doi: 10.1038/nature11532
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* 42, 961–967. doi: 10.1038/ng.695
- Huang, X., Yang, S., Gong, J., Zhao, Y., Feng, Q., Gong, H., et al. (2015). Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat. Commun.* 6:6258. doi: 10.1038/ncomms7258
- Huang, X., Zhao, Y., Wei, X., Li, C., Wang, A., Zhao, Q., 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. doi: 10.1038/ng.1018
- Huang, Y., Sun, C., Min, J., Chen, Y., Tong, C., and Bao, J. (2015). Association mapping of quantitative trait loci for mineral element contents in whole grain rice (*Oryza sativa* L.). *J. Agric. Food Chem.* 63, 10885–10892. doi: 10.1021/acs.jafc.5b04932
- Huo, H., Dahal, P., Kunusoth, K., McCallum, C. M., and Bradford, K. J. (2013). Expression of *9-cis-EPOXYCAROTENOID DIOXYGENASE4* is essential for thermoinhibition of lettuce seed germination but not for seed development or stress tolerance. *Plant Cell.* 25, 884–900. doi: 10.1105/tpc.112.108902
- Izawa, T., Konishi, S., Shomura, A., and Yano, M. (2009). DNA changes tell us about rice domestication. *Curr. Opin. Plant Biol.* 12, 185–192. doi: 10.1016/j.pbi.2009.01.004
- Jiao, Y., Wang, Y., Xue, D., Wang, J., Yan, M., Liu, G., et al. (2010). Regulation of *OsSPL14* by *OsmiR156* defines ideal plant architecture in rice. *Nat. Genet.* 42, 541–544. doi: 10.1038/ng.591
- Jin, J., Huang, W., Gao, J. P., Yang, J., Shi, M., Zhu, M. Z., et al. (2008). Genetic control of rice plant architecture under domestication. *Nat. Genet.* 40, 1365–1369. doi: 10.1038/ng.247
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723. doi: 10.1534/genetics.107.080101
- Kovach, M. J., Sweeney, M. T., and McCouch, S. R. (2007). New insights into the history of rice domestication. *Trends Genet.* 23, 578–587. doi: 10.1016/j.tig.2007.08.012
- Kulwal, P. L., Mir, R. R., Kumar, S., and Gupta, P. K. (2010). QTL analysis and molecular breeding for seed dormancy and pre-harvest sprouting tolerance in bread wheat. *J. Plant Biol.* 37, 59–74. Available online at: https://www.researchgate.net/publication/258927228_QTL_Analysis_and_Molecular

- Breeding_for_Seed_Dormancy_and_Pre-harvest_Sprouting_Tolerance_in_Bread_Wheat
- Kumar, V., Singh, A., Mithra, S. V. A., Krishnamurthy, S. L., Parida, S. K., Jain, S., et al. (2015). Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* 22, 133–145. doi: 10.1093/dnares/dsu046
- Lam, H. M., Xu, X., Liu, X., Chen, W., Yang, G., Wong, F. L., et al. (2010). Resequencing of 31 wild and cultivated soybean genomes identifies patterns of genetic diversity and selection. *Nat. Genet.* 42, 1053–1059. doi: 10.1038/ng.715
- Lee, G. A., Jeon, Y. A., Lee, H. S., Hyun, D. Y., Lee, J. R., Lee, M. C., et al. (2017). New genetic loci associated with preharvest sprouting and its evaluation based on the model equation in rice. *Front. Plant Sci.* 8:1393. doi: 10.3389/fpls.2017.01393
- Li, C., Ni, P., Francki, M., Hunter, A., Zhang, Y., Schibeci, D., et al. (2004). Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. *Funct. Integr. Genomics* 4, 84–93. doi: 10.1007/s10142-004-0104-3
- Li, C., Zhou, A., and Sang, T. (2006). Rice domestication by reducing shattering. *Science* 311, 1936–1939. doi: 10.1126/science.1123604
- Lijavetzky, D., Martínez, M. C., Carrari, F., and Hopp, H. E. (2000). QTL analysis and mapping of pre-harvest sprouting resistance in sorghum. *Euphytica* 112, 125–135. doi: 10.1023/A:1003823829878
- Liu, K., and Muse, S. V. (2005). PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics* 21, 2128–2129. doi: 10.1093/bioinformatics/bti282
- Liu, S., Sehgal, S. K., Li, J., Lin, M., Trick, H. N., Yu, J., et al. (2013). Cloning and characterization of a critical regulator for preharvest sprouting in wheat. *Genetics* 195, 263–273. doi: 10.1534/genetics.113.152330
- Liu, Y., Wang, L., Deng, M., Li, Z., Lu, Y., Wang, J., et al. (2015). Genome-wide association study of phosphorus-deficiency-tolerance traits in *Aegilops tauschii*. *Theor. Appl. Genet.* 128, 2203–2212. doi: 10.1007/s00122-015-2578-x
- Lu, Q., Zhang, M., Niu, X., Wang, C., Xu, Q., Feng, Y., et al. (2016). Uncovering novel loci for mesocotyl elongation and shoot length in *indica* rice through genome-wide association mapping. *Planta* 243, 645–657. doi: 10.1007/s00425-015-2434-x
- Lu, Q., Zhang, M., Niu, X., Wang, S., Xu, Q., Feng, Y., et al. (2015). Genetic variation and association mapping for 12 agronomic traits in *indica* rice. *BMC Genomics* 16:1067. doi: 10.1186/s12864-015-2245-2
- Luo, J., Liu, H., Zhou, T., Gu, B., Huang, X., Shanguan, Y., et al. (2013). *An-1* encodes a basic helix-loop-helix protein that regulates awn development, grain size, and grain number in rice. *Plant Cell* 25, 3360–3376. doi: 10.1105/tpc.113.113589
- Lv, Y., Guo, Z., Li, X., Ye, H., Li, X., and Xiong, L. (2015). New insights into the genetic basis of natural chilling and cold shock tolerance in rice by genome-wide association analysis. *Plant Cell Environ.* 39, 556–570. doi: 10.1111/pce.12635
- Magwa, R. A., Zhao, H., and Xing, Y. Z. (2016). Genome-wide association mapping revealed a diverse genetic basis of seed dormancy across subpopulations in rice (*Oryza sativa* L.). *BMC Genet.* 17:28. doi: 10.1186/s12863-016-0340-2
- Marzougui, S., Sugimoto, K., Yamanouchi, U., Shimono, M., Hoshino, T., Hori, K., et al. (2012). Mapping and characterization of seed dormancy QTLs using chromosome segment substitution lines in rice. *Theor. Appl. Genet.* 124, 893–902. doi: 10.1007/s00122-011-1753-y
- Masojć, P., Banek-Tabor, A., Milczarski, P., and Twardowska, M. (2007). QTLs for resistance to preharvest sprouting in rye (*Secale cereale* L.). *J. Appl. Genet.* 48, 211–217. doi: 10.1007/BF03195215
- Masojć, P., Lebiecka, K., Milczarski, P., Wiśniewska, M., Łań, A., and Owsianicki, R. (2009). Three classes of loci controlling preharvest sprouting in rye (*Secale cereale* L.) discerned by means of bidirectional selective genotyping (BSG). *Euphytica* 170:123. doi: 10.1007/s10681-009-9952-1
- Moonesinghe, R., Khoury, M. J., and Janssens, A. C. (2007). Most published research findings are false-but a little replication goes a long way. *PLoS Med.* 4:e28. doi: 10.1371/journal.pmed.0040028
- Neff, M. M., Turk, E., and Kalishman, M. (2002). Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet.* 18, 613–615. doi: 10.1016/S0168-9525(02)02820-2
- Nei, M. (1972). Genetic distance between populations. *Am. Nat.* 106, 283–292. doi: 10.1086/282771
- Olsen, K. M., Caicedo, A. L., Polato, N., McClung, A., McCouch, S., and Purugganan, M. D. (2006). Selection under domestication: evidence for a sweep in the rice waxy genomic region. *Genetics* 173, 975–983. doi: 10.1534/genetics.106.056473
- Pasam, R. K., Sharma, R., Walther, A., Özkan, H., Graner, A., and Kilian, B. (2014). Genetic diversity and population structure in a legacy collection of spring barley landraces adapted to a wide range of climates. *PLoS ONE* 9:e116164. doi: 10.1371/journal.pone.0116164
- Pearson, T. A., and Manolio, T. A. (2008). How to interpret a genome-wide association study. *JAMA* 299, 1335–1344. doi: 10.1001/jama.299.11.1335
- Pe'er, I., de Bakker, P. I., Maller, J., Yelensky, R., Altshuler, D., and Daly, M. J. (2006). Evaluating and improving power in whole-genome association studies using fixed marker sets. *Nat. Genet.* 38, 663–667. doi: 10.1038/ng1816
- Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., et al. (2014). Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genet.* 10:e1004148. doi: 10.1371/journal.pgen.1004148
- Qu, C. M., Li, S. M., Duan, X. J., Fan, J. H., Jia, J. D., Zhao, H. Y., et al. (2015). Identification of candidate genes for seed glucosinolate content using association mapping in *Brassica napus* L. *Genes* 6, 1215–1229. doi: 10.3390/genes6041215
- Remington, D. L., Thornsberry, J. M., Matsuoka, Y., Wilson, L. M., Whitt, S. R., Doebley, J., et al. (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc. Natl. Acad. Sci. U.S.A.* 98, 11479–11484. doi: 10.1073/pnas.201394398
- Richards, C. M., Brownson, M., Mitchell, S. E., Kresovich, S., and Panella, L. (2004). Polymorphic microsatellite markers for inferring diversity in wild and domesticated sugar beet (*Beta vulgaris*). *Mol. Ecol. Notes* 4, 243–245. doi: 10.1111/j.1471-8286.2004.00630.x
- Rohlf, F. J. (2000). *NTSYSpc: Numerical Taxonomy and Multivariate Analysis System, Version 2.1*. New York, NY: Exeter Software.
- Rostoks, N., Ramsay, L., MacKenzie, K., Cardle, L., Bhat, P. R., Roose, M. L., et al. (2006). Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proc. Natl. Acad. Sci. U.S.A.* 103, 18656–18661. doi: 10.1073/pnas.0606133103
- Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z., Richter, D. J., Schaffner, S. F., et al. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature* 419, 832–837. doi: 10.1038/nature01140
- Salvi, S., and Tuberosa, R. (2005). To clone or not to clone plant QTLs: present and future challenges. *Trends Plant Sci.* 10, 297–304.
- Sasaki, A., Ashikari, M., Ueguchi-Tanaka, M., Itoh, H., Nishimura, A., Datta, S. K., et al. (2002). A mutant gibberellin-synthesis gene in rice. *Nature* 416, 701–702. doi: 10.1038/416701a
- Schatzki, J., Schoo, B., Ecke, W., Herrfurth, C., Feussner, I., Becker, H. C., et al. (2013). Mapping of QTL for seed dormancy in a winter oilseed rape doubled haploid population. *Theor. Appl. Genet.* 126, 2405–2415. doi: 10.1007/s00122-013-2144-3
- Shewry, P. R., Napier, J. A., and Tatham, A. S. (1995). Seed storage proteins: structures and biosynthesis. *Plant Cell* 7, 945–956. doi: 10.1105/tpc.7.7.945
- Sugimoto, K., Takeuchi, Y., Ebana, K., Miyao, A., Hirochika, H., Hara, N., et al. (2010). Molecular cloning of *Sdr4*, a regulator involved in seed dormancy and domestication of rice. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5792–5797. doi: 10.1073/pnas.0911965107
- Sukumar, S., Dreisigacker, S., Lopes, M., Chavez, P., and Reynolds, M. P. (2015). Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor. Appl. Genet.* 128, 353–363. doi: 10.1007/s00122-014-2435-3
- Sweeney, M., and McCouch, S. (2007). The complex history of the domestication of rice. *Ann. Bot.* 100, 951–957. doi: 10.1093/aob/mcm128
- Tan, L., Li, X., Liu, F., Sun, X., Li, C., Zhu, Z., et al. (2008). Control of a key transition from prostrate to erect growth in rice domestication. *Nat. Genet.* 40, 1360–1364. doi: 10.1038/ng.197
- Teoh, K. T., Requesens, D. V., Devaiah, S. P., Johnson, D., Huang, X., Howard, J. A., et al. (2013). Transcriptome analysis of embryo maturation in maize. *BMC Plant Biol.* 13:19. doi: 10.1186/1471-2229-13-19
- Ullrich, S. E., Hayes, P. M., Dyer, W. E., Blake, T. K., and Clancy, J. A. (1993). “Quantitative trait locus analysis of seed dormancy in ‘Steptoe’ barley, in pre-harvest sprouting in cereals,” in *American Association of Cereal Chemists*, eds M. K. Walker-Simmons and J. L. Ried (St. Paul, MN: Academic), 136–145.

- Wan, J. M., Liang, L., Tang, J. Y., Wang, C. M., Hou, M. Y., Jing, W., et al. (2006). Genetic dissection of the seed dormancy trait in cultivated rice (*Oryza sativa* L.). *Plant Sci.* 170, 786–792. doi: 10.1016/j.plantsci.2005.11.011
- Wang, C., Yang, Y., Yuan, X., Xu, Q., Feng, Y., Yu, H., et al. (2014). Genome-wide association study of blast resistance in *indica* rice. *BMC Plant Biol.* 14:311. doi: 10.1186/s12870-014-0311-6
- Wang, C. H., Zheng, X. M., Xu, Q., Yuan, X. P., Huang, L., Zhou, H. F., et al. (2014). Genetic diversity and classification of *Oryza sativa* with emphasis on Chinese rice germplasm. *Heredity* 112, 489–496. doi: 10.1038/hdy.2013.130
- Wang, X., Jia, M. H., Ghai, P., Lee, F. N., and Jia, Y. (2015). Genome wide association of rice blast disease resistance and yield related components of rice. *Mol. Plant Microbe Interact.* 28, 1383–1392. doi: 10.1094/MPMI-06-15-0131-R
- Wei, X., Liu, K., Zhang, Y., Feng, Q., Wang, L., Zhao, Y., et al. (2015). Genetic discovery for oil production and quality in sesame. *Nat. Commun.* 6:8609. doi: 10.1038/ncomms9609
- Wen, Z., Tan, R., Yuan, J., Bales, C., Du, W., Zhang, S., et al. (2014). Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. *BMC Genomics* 15:809. doi: 10.1186/1471-2164-15-809
- White, C. N., and Rivin, C. J. (2000). Gibberellins and seed development in maize. II. Gibberellin synthesis inhibition enhances abscisic acid signaling in cultured embryos. *Plant Physiol.* 122, 1089–1097. doi: 10.1104/pp.122.4.1089
- Wilkinson, S., Lu, Z. H., Megens, H. J., Archibald, A. L., Haley, C., Jackson, I. J., et al. (2013). Signatures of diversifying selection in European pig breeds. *PLoS Genet.* 9:e1003453. doi: 10.1371/journal.pgen.1003453
- Xie, K., Jiang, L., Lu, B., Yang, C., Li, L., Liu, X., et al. (2011). Identification of QTLs for seed dormancy in rice (*Oryza sativa* L.). *Plant Breed.* 130, 328–332. doi: 10.1111/j.1439-0523.2010.01829.x
- Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., et al. (2011). A genome-wide search for signals of high-altitude adaptation in Tibetans. *Mol. Biol. Evol.* 28, 1003–1011. doi: 10.1093/molbev/msq277
- Yan, J., Yang, X., Shah, T., Sánchez-Villeda, H., Li, J., Warburton, M., et al. (2010). High-throughput SNP genotyping with the GoldenGate assay in maize. *Mol. Breed.* 25, 441–451. doi: 10.1007/s11032-009-9343-2
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi: 10.1038/ncomms6087
- Yang, X., Gao, S., Xu, Z., Zhang, Z., Prasanna, B. M., Li, L., et al. (2011). Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. *Mol. Breed.* 28, 511–526. doi: 10.1007/s11032-010-9500-7
- Ye, H., Feng, J., Zhang, L., Zhang, J., Mispan, M. S., Cao, Z., et al. (2015). Map-based cloning of seed dormancy1-2 identified a gibberellin synthesis gene regulating the development of endosperm-imposed dormancy in rice. *Plant Physiol.* 169, 2152–2165. doi: 10.1104/pp.15.01202
- Yu, J., Pressoir, G., Briggs, W. H., Vroh Bi, I., Yamasaki, M., Doebley, J. F., et al. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zhang, J., Song, Q., Cregan, P. B., Nelson, R. L., Wang, X., Wu, J., et al. (2015). Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics.* 16:217. doi: 10.1186/s12864-015-1441-4
- Zhang, D., Kong, W., Robertson, J., Goff, V. H., Epps, E., Kerr, A., et al. (2015). Genetic analysis of inflorescence and plant height components in sorghum (*Panicoidae*) and comparative genetics with rice (*Oryzoidae*). *BMC Plant Biol.* 15:107. doi: 10.1186/s12870-015-0477-6
- Zhang, H., Tan, G., Yang, L., Yang, J., Zhang, J., and Zhao, B. (2009). Hormones in the grains and roots in relation to post-anthesis development of inferior and superior spikelets in *japonica / indica* hybrid rice. *Plant Physiol. Biochem.* 47, 195–204. doi: 10.1016/j.plaphy.2008.11.012
- Zhang, Z., Buckler, E. S., Casstevens, T. M., and Bradbury, P. J. (2009). Software engineering the mixed model for genome-wide association studies on large samples. *Brief. Bioinformatics.* 10, 664–675. doi: 10.1093/bib/bbp050
- Zhang, J., Zhao, J., Liang, Y., and Zou, Z. (2016a). Genome-wide association-mapping for fruit quality traits in tomato. *Euphytica* 207, 439–451. doi: 10.1007/s10681-015-1567-0
- Zhang, J., Song, Q., Cregan, P. B., and Jiang, G. L. (2016b). Genome-wide association study, genomic prediction and marker-assisted selection for seed weight in soybean (*Glycine max*). *Theor. Appl. Genet.* 129, 117–130. doi: 10.1007/s00122-015-2614-x
- Zhang, N., Gibon, Y., Lepak, N., Li, P., Dedow, L., Chen, C., et al. (2014). Genome-wide association of carbon and nitrogen metabolism in the maize nested association mapping population. *bioRxiv.* 2014. doi: 10.1101/010785
- Zhang, Z., Ersoz, E., Lai, C. Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., et al. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546
- Zhu, B. F., Si, L., Wang, Z., Zhou, Y., Zhu, J., Shangguan, Y., et al. (2011). Genetic control of a transition from black to straw-white seed hull in rice domestication. *Plant Physiol.* 155, 1301–1311. doi: 10.1104/pp.110.168500

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

Copyright © 2018 Lu, Niu, Zhang, Wang, Xu, Feng, Yang, Wang, Yuan, Yu, Wang, Chen, Liang and Wei. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.