Integrating GWAS and Transcriptomics to Identify Genes Involved in Seed Dormancy in Rice

Jin Shi

Shanghai Jiao Tong University School of Life Sciences and Biotechnology

Jianxin Shi

Shanghai Jiao Tong University School of Life Sciences and Biotechnology

Wanqi Liang

Shanghai Jiao Tong University School of Life Sciences and Biotechnology

Dabing Zhang (✉ zhangdb@sjtu.edu.cn)

Shanghai Jiao Tong University School of Life Sciences and Biotechnology <https://orcid.org/0000-0003-3181-9812>

Research Article

Keywords: Genome-wide association study, pre-harvest sprouting, rice, RNA-seq, seed dormancy

Posted Date: April 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-414570/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on July 26th, 2021. See the published version at <https://doi.org/10.1007/s00122-021-03911-1>.

Abstract

Seed dormancy is a key agricultural trait to prevent pre-harvest sprouting in crop plants such as rice (*Oryza sativa* L.), wheat (*Triticum aestivum*), and barley (*Hordeum vulgare* L.). However, our knowledge of seed dormancy is hampered by the complexities of studying a trait that changes over time after seed harvest, and is complicated by interactions between phytohormones, other regulators, and the environment. Here, we have conducted a genome-wide association study using a panel of 311 natural accessions of cultivated rice, examining a total of 519,158 single nucleotide polymorphisms (SNPs). Eight quantitative trait loci (QTLs) were found to associate with seed dormancy; expression of candidate genes within 100 kb of each QTL was examined in two published, germination-specific transcriptomic datasets. Ten candidate genes, differentially expressed within the first four days post-imbibition, were identified. Five of these genes had previously been associated with awn length, heading date, yield, and spikelet length phenotypes. In addition, previously identified genes involved in hormone signaling during germination were found to be differentially expressed between a japonica and an indica line; SNPs in the promoter of *Os9BGlu33* were associated with germination index. Collectively, our results are useful for future characterization of seed dormancy mechanism and crop improvement, and suggest haplotypes for further analysis that may be of use to boost PHS resistance in rice.

Key Messages

Several QTLs and genes responsible for seed dormancy were detected and SNP candidates were shown to cause changes in seed germination.

Introduction

Rice (*Oryza sativa*) is the staple food for over half the world's population, so safeguarding its production in the face of global warming is critical for food security. Changing environmental conditions can severely affect rice growth and yield; for example, prolonged rainfall and high humidity during rice maturation can lead to seed germination in rice panicles prior to harvest, giving rise to heavy economic and production losses (Zhu et al. 2019). This phenomenon is termed *pre-harvest sprouting* (PHS) and can occur due to low seed dormancy. Further understanding of seed dormancy is required to avoid PHS in different environments.

Seed dormancy is “one of the least understood phenomena in seed biology” (Finkelstein et al. 2008). Medium seed dormancy is regarded as a desirable domestication trait (Lohwasser et al. 2005), as it can ensure uniform germination of crops and prevent the risk of PHS (Gubler et al. 2005). Dormancy is known to be affected by the balance between two phytohormones, gibberellic acid (GA) and abscisic acid (ABA), but they are not the sole regulators of dormancy (Finkelstein et al. 2008). Other factors, such as seed condition, seed coat color, and environmental signals all influence dormancy (Furukawa et al. 2007; Sweeney et al. 2006). Many quantitative trait loci (QTLs) for seed dormancy have been identified in crops and some model plants, such as *Arabidopsis* (Atwell et al. 2010; Basbous-Serhal et al. 2016), soybean

(Dougherty and Boerma 1984), barley (Gao et al. 2003; Gawenda et al. 2015; Vanhala and Stam 2006), wheat (Martinez et al. 2018; Zhang et al. 2019; Zhu et al. 2019), sorghum (Cantoro et al. 2016; Lijavetzky et al. 2000), and rice (Gu et al. 2010; Gu et al. 2004; Gu et al. 2005; Sugimoto et al. 2010), but only four seed dormancy QTLs have been fine-cloned in rice using traditional mapping populations (Gu et al. 2011; Gu et al. 2010; Gu et al. 2004; Sugimoto et al. 2010). Due to the limited recombination events in inbred rice line populations, the power of the traditional mapping using bi-parental populations is limited.

Genome-wide association studies (GWAS) have been applied in plants to accelerate crop improvement and QTL study (Atwell et al. 2010; Huang et al. 2010; Yang et al. 2014). Using GWAS to study rice dormancy has been performed by several groups using germination rate and index as the key phenotypic traits, but due to differences in populations and data processing, different genes and QTLs have been identified: Lu et al. (2018) found *Sdr4* and qSD7-1 to be significant (Lu et al. 2018), while Magwa *et al.* (2016) found that *GAox3* and *Pid3* associate with dormancy (Magwa et al. 2016). These studies show great power and credence, but not all SNPs and loci mapped can be resolved as not all genes in the genetic maps have been annotated, so candidate genes are difficult to select. More information, and cross-validation from different techniques and studies, will help improve the accuracy of selecting candidate genes associated with dormancy.

Here, we use GWAS on a new rice association panel paired with existing transcriptomic datasets from the literature to identify candidate genes and SNPs affecting gene expression (eSNPs) in different cultivars. Candidate genes containing SNPs identified from GWAS are checked for their differential expression in two different RNA-seq datasets. Relationships with associated traits are determined by cross-checking SNPs in candidate gene with the RiceVarMap database (Zhao et al. 2015), and eSNPs for previously reported genes are assessed for effect on germination. Our results provide a basis for further elucidating the molecular mechanism behind PHS and seed dormancy of rice.

Materials And Methods

Plant materials

The association mapping population consisted of 311 rice accessions in three subgroups (japonica, indica, and mixed) as previously described (Wang et al. 2018). These cultivars were grown in paddy fields around Shanghai, China in May 2018. Cultivars with phenotypes of interest were re-planted in May 2019. Each accession was planted in one row with 6 hills per row, and grown to maturity in the field.

The heading time of each accession was recorded when the first panicle emerged. To minimize the influence of different heading time, spikes from different plants were harvested in September (2018 or 2019), approximately 30–40 days after heading.

Phenotype assessment for seed dormancy

After harvest, seeds were immediately removed from the spike and sterilized using 0.6 % (v/v) sodium hypochlorite solution for 15 min, and rinsed with distilled water. Two replicates of 30 seeds were placed on two layers of Whatman #1 filter paper in 9 cm Petri dishes, wet with 20 ml sterile water. The Petri dishes were incubated for six days at 28 °C / 100 % humidity in the light condition.

Germination was defined as emergence of the rootlet. From the second day after imbibition, germinated seeds were counted and removed daily. Germination index (GI) on day 6 was determined as the proportion of seeds (out of the original 30) that had germinated. GI results are presented as the mean of two replicates (Online Resource 1).

Association mapping

SNP data for our population was downloaded from Rice SNP-Seek Database (<https://snp-seek.irri.org/>, core dataset) (Mansueto et al. 2017). Accessions in our population were filtered using plink (v1.9) *-keep* command (Purcell et al. 2007), and SNPs further filtered using plink with maf (0.05) and geno (0.2). Population structure was analyzed in several steps, first using plink (v1.9) *-ibs-matrix* command (Purcell et al. 2007), and the tree file generated by phylip (v3.6) using neighbour.exe (Retief 2000). The phylogenetic tree was visualized in Figtree (v1.4.4), available through the github rambaut/figtree repository. The plink files were used to calculate the indicator p value using Genetic type 1 Error Calculator (GEC) software (Duggal et al. 2008), which gave 6.67E-6 as the suggestive p value ($-\log(p\text{-value})=5.175$).

We applied two models for the GWAS analysis: Factored Spectrally Transformed Linear Mixed Model (fastLMM) was conducted using FastLmm (v2.07) (Lippert et al. 2014); and Multi-Locus Mixed Model (MLMM) was conducted using TASSEL for format conversion and GAPIT package in R (Bradbury et al. 2007; Tang et al. 2016). Manhattan plots were generated using ggplot2 package in R (Villanueva and Chen 2019).

Fixation index analysis

The fixation index between subgroups were calculated using vcftools (v0.1.13) (Danecek et al. 2011). The window size was set as 20kb and window step set as 10kb. Fixation index plot was made using pheatmap package in R.

Principal component analysis

The principal component analysis was conducted using plink (v1.9) *-pca* command (Purcell et al. 2007). The first two principal components were extracted. PCA plot was made in R with subpopulations assigned with different colors.

Transcriptome analysis

Two rice transcriptome datasets (SRP277875, GSE115371) were downloaded from the Gene Expression Omnibus (*GEO*) database using the *Sequence Read Archive* (*SRA*) toolkit (Narsai et al. 2017; Yang et al. 2020). For GSE115371 dataset, only timepoints from aerobically grown samples were used, which includes 0 h, 1 h, 3 h, 12 h, 1 d, 2 d, 3 d and 4 d timepoints after imbibition. The raw data was quality controlled using Trimmomatic (v0.39) software, and mapped to MSU 7.0 rice reference genome using Hisat2 (v2.1.0) (Bolger et al. 2014; Kawahara et al. 2013; Kim et al. 2015). The aligned files were sorted and turned into bam files using samtools (v1.9) (Li et al. 2009). Reads aligned were counted using HTSeq package (Anders et al. 2015). For differential expression analysis, R package DeSeq2 was used. Genes with a false discovery rate < 0.05 and $|\log_2(\text{fold change})| > 1$ were considered as DEGs (Love et al. 2014). RPKM data was calculated using read counts and heatmap plots were made based on the RPKM data using pheatmap package in R.

Haplotype analysis

SNP data for the candidate genes were extracted from the obtained SNP dataset. This dataset only contained biallelic SNPs. Heterozygous and missing alleles are excluded from further haplotype analysis. For genes found in QTLs, only non-synonymous SNPs within the coding regions of these genes were applied for haplotype analysis, while for other five genes, which have been previously identified to be involved in hormone signaling during germination, SNPs within the promotor region (3kb upstream of the gene) were used for haplotype analysis to find eSNPs. Haplotype analysis was conducted in R to determine whether the loci can cause germination index change by Student's *t*-test. The results were visualized using GraphPad Prism (v8.4.2). The number of accessions correspond to each SNPs was calculated and checked manually.

Phylogenetics of the AIRP genes

The homologs of LOC_Os07g17400 in rice (*Oryza Sativa*), sorghum (*Sorghum bicolor*) and Arabidopsis (*Arabidopsis thaliana*) were extracted from SALAD database (v3) (Mihara et al. 2010). To further include sequences from barley (*Hordeum vulgare*) and wheat (*Triticum aestivum*), the sequence of LOC_Os07g17400 was used for blast search on NCBI. All of the obtained sequences were aligned using MUSCLE algorithm (Edgar 2004). MEGA6 software was used to create a phylogenetic tree. Neighbour-

joining methods were used with a p -distance model and pairwise detection and bootstrap with 1000 replicates (Tamura et al. 2013).

Results

Population structure and germination differences

Analysis of SNPs in an association panel of 311 natural rice accessions using phylogenetics and fixation index analysis revealed that the japonica ($n=160$) and indica ($n=138$) varieties are easily differentiated (Fig. 1). The mixed subgroup ($n=13$), comprising mainly Australian and Basmati cultivars, had more members similar to the indica subpopulation.

The germination index (GI) of harvested seeds was used to test the degree of seed dormancy for each accession. The GI average across all 311 lines was 0.24, with the japonica GI slightly higher, at 0.25, while the indica GI was 0.23, with a higher standard deviation between accessions (Table 1). The mixed group had the lowest GI (0.15): while their group was too small to draw any general conclusions, their inclusion in the subsequent GWAS analysis may provide insights into the mechanisms of seed dormancy and pre-harvest sprouting.

GWAS and candidate gene selection

A total of 519,158 SNPs were included for association analysis. We used two different models for GWAS, fastLMM and MLMM (Fig. 2). The QTLs were assigned when the peak SNP correlation exceeded the calculated threshold. Eight QTLs, each of 200 kb, were identified (Fig. 2; Table 2).

Expression of candidate genes within those QTLs were examined using two transcriptomic datasets from the GEO database. Dataset GSE115371 provided gene expression data at 0 h, 1 h, 3 h, 12 h, 1 d, 2 d, 3 d and 4 d timepoints after imbibition, for an Australian variety (Amaroo); only timepoints from aerobically grown samples were used (Narsai et al. 2017). DEGs were identified by comparing expression at different timepoints to expression at 0 h. Dataset SRP277875 provided expression data for a japonica variety (02428) and an indica variety (YZX) at 2, 3, and 4 days after imbibition compared with a 0 h timepoint (Yang et al. 2020). The data of the japonica variety was used for expression comparison of candidate genes (Fig. 3a).

Ten candidate genes from five QTLs identified by GWAS were differentially expressed across the first four days post-imbibition (Table 2; Fig. 3). Five genes had similar expression patterns in both datasets (Fig. 3), namely LOC_Os07g17400, *OsNFYA5* (expressed at 0 h and soon after imbibition); LOC_Os08g06090, *PPT1* (2 d post-imbibition); and *ROMT9* (3 d post-imbibition). Two further genes, *OsYAB1* and *OsFLA16*, had similar patterns of expression in the two datasets, that peaked at 2 or 3 d post-imbibition. A further three genes — *OsPUB6*, LOC_Os09g12650, and LOC_Os09g21370 — has quite distinct patterns of expression, possibly due to varietal differences.

Bioinformatic analysis of the candidates

SNP variations in five candidate genes had previously been associated with phenotypes such as awn length, heading date, yield, and spikelet length (RiceVarMap; Online Resource 3). Three candidates, LOC_Os07g17400, *PPT1*, and *OsFLA16*, are reported to associate with awn length.

Haplotype analysis confirmed that three SNPs in our candidate genes associated with GI caused non-synonymous mutations (Online Resource 4). In *ROMT9*, a glycine to serine change caused GI to drop. Similarly, a threonine to alanine change in *PPT1* reduced GI, but the number of T-haplotype varieties was very small so further verification is required. An asparagine to aspartate change in LOC_Os09g12650 caused GI to increase. Phylogenetic analysis of the predicted protein encoded by LOC_Os07g17400 reveals homology to AIRP family proteins in maize, *Triticum*, and Arabidopsis (Online Resource 6).

eSNPs associated with seed germination

To further find eSNPs associated with seed germination, we compared expression in the japonica and indica varieties in dataset SRP277875 (Yang et al. 2020). More than twice as many genes were up-regulated in the japonica than in the indica variety (Online Resource 5), suggesting that the germination difference between these two rice subpopulations can be caused by the differential expression of genes.

Five genes known to be involved in seed germination were differentially expressed in this dataset (Table 3). *OsGA20ox1*, *OsGAP*, and *Os9BGl33* were up-regulated in japonica 02428, while *OsGA3ox2* and *OsPP2C51* were up-regulated in indica YZX. The timing and extent of differential expression differed for each gene. The largest change was in *OsGAP*, whose expression was 6.5× higher in the japonica variety at 0 h, and 2.3× higher at the 3 d timepoint. We found four mutations in the promoter of *Os9BGl33* that are associated with a change in GI (Fig. 4; Online Resource 1).

Discussion

Candidate genes for seed dormancy

Seed dormancy is not a very stable trait, as the ability of seeds to break dormancy increases over time. Here, we have combined GWAS with transcriptomics analyses to cross-validate phenotyping results. Three out of ten candidate genes associated with dormancy had previously been linked with awn length (Online Resource 3); as awns tend to reduce germination rate (Ntakirutimana et al. 2019), we suggest that these genes may regulate seed germination by regulating awn development in natural populations. As long awns were important for seed dispersal in wild rice and gradually lost during domestication (Luo et al. 2013), awn loss may cause loss of desired dormancy in domesticated lines. Thus, GWAS from natural populations can reveal mechanisms that may be lost using bi-parental mapping on inbred lines, which in this example, may have been selected for short or missing awns.

For the awn-associated candidate LOC_Os09g12650, its homolog in Arabidopsis, AtAIRP2, is a cytosolic RING-type E3 ubiquitin ligase that can positively regulate ABA response in Arabidopsis, while its knockout mutant is insensitive to ABA during seed germination (Oh et al. 2017). The involvement of LOC_Os07g17400 in ABA response is consistent with its proposed role in rice seed dormancy.

Differential expression of seed dormancy-related genes in japonica and indica rice

The GI of japonica rice was generally higher than for indica rice varieties (Table 1), and nearly twice as many genes were up-regulated in the 4 days post-imbibition in japonica 02428 than were up-regulated in indica YZX (Online Resource 5). To examine whether these genomic variations affected the expression of seed dormancy-related genes, we examined five DEGs previously reported to be involved in seed germination (Table 3). *OsGA20ox1* and *OsGA3ox2* are genes involved in GA biosynthesis (Kaneko et al. 2002; Oikawa et al. 2004), whose expression would affect overall levels of GA, influencing seed dormancy (Guo et al. 2013). OsGAP is a GTPase-activating protein that can integrate ABA and reactive oxygen species (ROS) signaling (Xu et al. 2019); its knockout leads to ABA and H₂O₂ accumulation (Xu et al. 2019). OsPP2C51 is a member of the phosphatase 2C clade A protein family, which can interact with PYL family to transmit ABA signals to positively regulate seed dormancy (Bhatnagar et al. 2017). *OsPP2CC51* is down-regulated in japonica 02428, consistent with our results that japonica lines have a higher GI. Mutants of *Os9BGlu33* exhibit a lower germination rate than wild type seeds, but which are less inhibited than wild type in the presence of ABA, a germination inhibitor (Ren et al. 2019).

Here, we have paired GWAS and transcriptomics analyses to identify eight QTLs associated with seed dormancy, and selected ten candidate genes within five of those QTLs that exhibited differential expression patterns during germination (Table 2). Further, five genes related to hormone-responsive seed dormancy were found to be differentially expressed between japonica and indica varieties, and SNPs in the promoter of one of them, *Os9BGlu33*, were shown to cause changes in GI. Further experiments to validate these findings will lead to new haplotypes for inclusion into breeding programs to generate cultivars with stronger resistance to PHS.

Declarations

Acknowledgements

We would like to thank Mr. Zhijing Luo and Ms. Mingjiao Chen for their assistance in rice growth, and Dr. Natalie Betts for her valuable suggestions on the article.

Funding

Not applicable.

Conflict of interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Availability of data and material

The association mapping population data can be found in Rice SNP-Seek database (Wang et al. 2018), and transcriptome data can be found in the Gene Expression Omnibus (*GEO*) database, with the accession number SRP277875 and GSE115371 (Narsai et al. 2017; Yang et al. 2020).

Code availability

Not applicable.

Authors' contributions

Dabing Zhang and Wanqi Liang designed the study. Jin Shi performed the experiments and bioinformatic analysis. Jianxin Shi helped to revise the manuscript. All authors contributed to the article and approved the submitted version.

References

1. Anders S, Pyl PT, Huber W (2015) HTSeq-a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31:166–169
2. Atwell S, Huang YS, Vilhjalmsdottir BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Mulyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JDG, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010) Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465:627–631
3. Basbous-Serhal I, Leymarie J, Bailly C (2016) Fluctuation of *Arabidopsis* seed dormancy with relative humidity and temperature during dry storage. *J Exp Bot* 67:119–130
4. Bhatnagar N, Min M-K, Choi E-H, Kim N, Moon S-J, Yoon I, Kwon T, Jung K-H, Kim B-G (2017) The protein phosphatase 2C clade A protein OsPP2C51 positively regulates seed germination by directly inactivating OsbZIP10. *Plant Mol Biol* 93:389–401
5. Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
6. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635

7. Cantoro R, Fernandez LG, Cervigni GDL, Rodriguez MV, Gieco JO, Paniego N, Heinz RA, Benech-Arnold RL (2016) Seed dormancy QTL identification across a *Sorghum bicolor* segregating population. *Euphytica* 211:41–56
8. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Anal G (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
9. Dougherty RW, Boerma HR (1984) GENOTYPIC VARIATION FOR RESISTANCE TO PREHARVEST SPROUTING IN SOYBEAN. *Crop Sci* 24:683–686
10. Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE (2008) Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *Bmc Genomics* 9
11. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
12. Finkelstein R, Reeves W, Ariizumi T, Steber C (2008) Molecular aspects of seed dormancy. *Annu Rev Plant Biol* 59:387–415
13. Furukawa T, Maekawa M, Oki T, Suda I, Iida S, Shimada H, Takamure I, Kadowaki K-i (2007) The Rc and Rd genes are involved in proanthocyanidin synthesis in rice pericarp. *Plant J* 49:91–102
14. Gao W, Clancy JA, Han F, Prada D, Kleinhofs A, Ullrich SE (2003) Molecular dissection of a dormancy QTL region near the chromosome 7 (5H) L telomere in barley. *Theor Appl Genet* 107:552–559
15. Gawenda I, Thorwarth P, Guenther T, Ordon F, Schmid KJ (2015) Genome-wide association studies in elite varieties of German winter barley using single-marker and haplotype-based methods. *Plant Breeding* 134:28–39
16. Gu X-Y, Foley ME, Horvath DP, Anderson JV, Feng J, Zhang L, Mowry CR, Ye H, Suttle JC, Kadowaki K-i, Chen Z (2011) Association Between Seed Dormancy and Pericarp Color Is Controlled by a Pleiotropic Gene That Regulates Absciscic Acid and Flavonoid Synthesis in Weedy Red Rice. *Genetics* 189:1515–1515+
17. Gu X-Y, Liu T, Feng J, Suttle JC, Gibbons J (2010) The qSD12 underlying gene promotes absciscic acid accumulation in early developing seeds to induce primary dormancy in rice. *Plant Mol Biol* 73:97–104
18. Gu XY, Kianian SF, Foley ME (2004) Multiple loci and epistases control genetic variation for seed dormancy in weedy rice (*Oryza sativa*). *Genetics* 166:1503–1516
19. Gu XY, Kianian SF, Hareland GA, Hoffer BL, Foley ME (2005) Genetic analysis of adaptive syndromes interrelated with seed dormancy in weedy rice (*Oryza sativa*). *Theor Appl Genet* 110:1108–1118
20. Gubler F, Millar AA, Jacobsen JV (2005) Dormancy release, ABA and pre-harvest sprouting. *Curr Opin Plant Biol* 8:183–187
21. Guo X, Hou X, Fang J, Wei P, Xu B, Chen M, Feng Y, Chu C (2013) The rice GERMINATION DEFECTIVE 1, encoding a B3 domain transcriptional repressor, regulates seed germination and seedling development by integrating GA and carbohydrate metabolism. *Plant J* 75:403–416

22. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q-F, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–976
23. Kaneko M, Itoh H, Ueguchi-Tanaka M, Ashikari M, Matsuoka M (2002) The alpha-amylase induction in endosperm during rice seed germination is caused by gibberellin synthesized in epithelium. *Plant Physiol* 128:1264–1270
24. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL, Davidson RM, Lin H, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR, Matsumoto T (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6
25. Kim D, Langmead B, Salzberg SL (2015) HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* 12:357–121
26. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data P (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079
27. Lijavetzky D, Martinez MC, Carrari F, Hopp HE (2000) QTL analysis and mapping of pre-harvest sprouting resistance in sorghum. *Euphytica* 112:125–135
28. Lippert C, Xiang J, Horta D, Widmer C, Kadie C, Heckerman D, Listgarten J (2014) Greater power and computational efficiency for kernel-based association testing of sets of genetic variants. *Bioinformatics* 30:3206–3214
29. Lohwasser U, Roder MS, Borner A (2005) QTL mapping of the domestication traits pre-harvest sprouting and dormancy in wheat (*Triticum aestivum* L.). *Euphytica* 143:247–249
30. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* 15
31. 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, Wei X (2018) Genome-Wide Association Study of Seed Dormancy and the Genomic Consequences of Improvement Footprints in Rice (*Oryza sativa* L.). *Frontiers in Plant Science* 8
32. Luo J, Liu H, Zhou T, Gu B, Huang X, Shangguan Y, Zhu J, Li Y, Zhao Y, Wang Y, Zhao Q, Wang A, Wang Z, Sang T, Wang Z, Han B (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
33. Magwa RA, Zhao H, Xing Y (2016) Genome-wide association mapping revealed a diverse genetic basis of seed dormancy across subpopulations in rice (*Oryza sativa* L.). *Bmc Genetics* 17
34. Mansueto L, Rommel Fuentes R, Nikki Borja F, Detras J, Miguel Abriol-Santos J, Chebotarov D, Sanciangco M, Palis K, Copetti D, Poliakov A, Dubchak I, Solovyev V, Wing RA, Sackville Hamilton R, Mauleon R, McNally KL, Alexandrov N (2017) Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Res* 45:D1075–D1081

35. Martinez SA, Godoy J, Huang M, Zhang Z, Carter AH, Campbell KAG, Steber CM (2018) Genome-Wide Association Mapping for Tolerance to Preharvest Sprouting and Low Falling Numbers in Wheat. *Frontiers in Plant Science* 9
36. Mihara M, Itoh T, Izawa T (2010) SALAD database: a motif-based database of protein annotations for plant comparative genomics. *Nucleic Acids Res* 38:D835–D842
37. Narsai R, Secco D, Schultz MD, Ecker JR, Lister R, Whelan J (2017) Dynamic and rapid changes in the transcriptome and epigenome during germination and in developing rice (*Oryza sativa*) coleoptiles under anoxia and re-oxygenation. *Plant J* 89:805–824
38. Ntakirutimana F, Xiao B, Xie W, Zhang J, Zhang Z, Wang N, Yan J (2019) Potential Effects of Awn Length Variation on Seed Yield and Components, Seed Dispersal and Germination Performance in Siberian Wildrye (*Elymus sibiricus* L.). *Plants-Basel* 8
39. Oh TR, Kim JH, Cho SK, Ryu MY, Yang SW, Kim WT (2017) AtAIRP2 E3 Ligase Affects ABA and High-Salinity Responses by Stimulating Its ATP1/SDIRIP1 Substrate Turnover(1). *Plant Physiol* 174:2515–2531
40. Oikawa T, Koshioka M, Kojima K, Yoshida H, Kawata M (2004) A role of OsGA20ox1, encoding an isoform of gibberellin 20-oxidase, for regulation of plant stature in rice. *Plant Mol Biol* 55:687–700
41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
42. Ren R, Li D, Zhen C, Chen D, Chen X (2019) Specific roles of Os4BGlu10, Os6BGlu24, and Os9BGlu33 in seed germination, root elongation, and drought tolerance in rice. *Planta* 249:1851–1861
43. Retief JD (2000) Phylogenetic analysis using PHYLIP. *Methods in molecular biology*. (Clifton NJ) 132:243–258
44. Sugimoto K, Takeuchi Y, Ebana K, Miyao A, Hirochika H, Hara N, Ishiyama K, Kobayashi M, Ban Y, Hattori T, Yano M (2010) Molecular cloning of Sdr4, a regulator involved in seed dormancy and domestication of rice. *Proc Natl Acad Sci USA* 107:5792–5797
45. Sweeney MT, Thomson MJ, Pfeil BE, McCouch S (2006) Caught red-handed: Rc encodes a basic helix-loop-helix protein conditioning red pericarp in rice. *Plant Cell* 18:283–294
46. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Mol Biol Evol* 30:2725–2729
47. Tang Y, Liu X, Wang J, Li M, Wang Q, Tian F, Su Z, Pan Y, Liu D, Lipka AE, Buckler ES, Zhang Z (2016) GAPIT Version 2: An Enhanced Integrated Tool for Genomic Association and Prediction. *Plant Genome* 9
48. Vanhala TK, Stam P (2006) Quantitative trait loci for seed dormancy in wild barley (*Hordeum spontaneum* c. Koch). *Genet Resour Crop Evol* 53:1013–1019
49. Villanueva RAM, Chen ZJ (2019) ggplot2: Elegant Graphics for Data Analysis, 2nd edition. *Measurement-Interdisciplinary Research and Perspectives* 17:160–167

50. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, Mansueto L, Copetti D, Sanciangco M, Palis KC, Xu J, Sun C, Fu B, Zhang H, Gao Y, Zhao X, Shen F, Cui X, Yu H, Li Z, Chen M, Detras J, Zhou Y, Zhang X, Zhao Y, Kudrna D, Wang C, Li R, Jia B, Lu J, He X, Dong Z, Xu J, Li Y, Wang M, Shi J, Li J, Zhang D, Lee S, Hu W, Poliakov A, Dubchak I, Ulat VJ, Borja FN, Mendoza JR, Ali J, Gao Q, Niu Y, Yue Z, Naredo MEB, Talag J, Wang X, Li J, Fang X, Yin Y, Glaszmann J-C, Zhang J, Li J, Hamilton RS, Wing RA, Ruan J, Zhang G, Wei C, Alexandrov N, McNally KL, Li Z, Leung H (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557:43–43+
51. Xu F, Tang J, Gao S, Chang X, Du L, Chu C (2019) Control of rice pre-harvest sprouting by glutaredoxin-mediated abscisic acid signaling. *Plant Journal*
52. Yang J, Su L, Li DD, Luo LX, Sun K, Yang M, Gu FW, Xia AY, Liu YZ, Wang H, Chen ZQ, Guo T (2020) Dynamic transcriptome and metabolome analyses of two types of rice during the seed germination and young seedling growth stages. *Bmc Genomics* 21
53. Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, Fang W, Feng H, Xie W, Lian X, Wang G, Luo Q, Zhang Q, Liu Q, Xiong L (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nature Communications* 5
54. Zhang W, Zhao J, He J, Wang X, Kang L, Yuan H, Chen D (2019) Identification and Validation of Five Molecular Markers for Pre-harvest Sprouting Tolerance in Ningxia Wheat Varieties(Lines). *Journal of Triticeae Crops* 39:532–539
55. Zhao H, Yao W, Ouyang Y, Yang W, Wang G, Lian X, Xing Y, Chen L, Xie W (2015) RiceVarMap: a comprehensive database of rice genomic variations. *Nucleic Acids Res* 43:D1018–D1022
56. Zhu Y, Wang S, Wei W, Xie H, Liu K, Zhang C, Wu Z, Jiang H, Cao J, Zhao L, Lu J, Zhang H, Chang C, Xia X, Xiao S, Ma C (2019) Genome-wide association study of pre-harvest sprouting tolerance using a 90K SNP array in common wheat (*Triticum aestivum* L.). *Theor Appl Genet* 132:2947–2963

Tables

Table 1. Germination index statistics in different rice populations

Population	Accessions (#)	Germination Index	SD	Min	Median	Max
All	311	0.24	0.19	0	0.19	0.76
Japonica	160	0.25	0.16	0	0.22	0.67
Indica	138	0.23	0.22	0	0.14	0.76
Mixed	13	0.15	0.17	0	0.08	0.59

Table 2. QTL information and candidate genes in the region

QTL name	Position of peak SNP	Model	Candidate genes	Description
2018QTL-1	Chr2: 26400956	fastLMM		
2018QTL-2	Chr4: 12919091	MLMM		
2018QTL-3	Chr7: 3171375	MLMM	LOC_Os07g06470 (<i>OsNFYA5</i>)	nuclear transcription factor Y subunit
			LOC_Os07g06620 (<i>OsYAB1</i>)	YABBY domain containing protein
			LOC_Os07g06680 (<i>OsFLA16</i>)	fasciclin domain containing protein
2018QTL-4	Chr7: 10291730	fastLMM	LOC_Os07g17400	zinc finger, RING-type domain containing protein
2018QTL-5	Chr8: 3324427	MLMM	LOC_Os08g06090	zinc finger, C3HC4-type domain containing protein
			LOC_Os08g06100 (<i>ROMT9</i>)	O-methyltransferase
2018QTL-6	Chr9: 7212442	MLMM	LOC_Os09g12600 (<i>PPT1</i>)	phosphate/phosphate translocator
			LOC_Os09g12650	pyrophosphate-fructose 6-phosphate 1-phosphotransferase subunit alpha
2018QTL-7	Chr9: 12832194	MLMM	LOC_Os09g21120 (<i>OsPUB6</i>)	armadillo/beta-catenin repeat family protein
			LOC_Os09g21370	cysteine proteinase EP-B 1 precursor
2018QTL-8	Chr11: 16712860	fastLMM		

Table 3. Differentially expressed gene between japonica and indica sub-species

Gene	Gene ID	Expression in japonica vs indica	Differential expression period	Log ₂ (fold change)
<i>OsGA20ox1</i>	LOC_Os03g63970	Up-regulated	0 h	2.64
<i>OsGAP</i>	LOC_Os07g31720	Up-regulated	0 h/3 d	2.70/1.23
<i>OsGA3ox2</i>	LOC_Os01g08220	Down-regulated	2 d	1.83
<i>OsPP2C51</i>	LOC_Os05g49730	Down-regulated	2 d	1.73
<i>Os9BGl33</i>	LOC_Os09g33710	Up-regulated	3 d	1.90

Figures

Figure 1

Population structure of 311 accessions of rice. a Neighbor-joining tree based on 311 samples and computed on a simple matching distance matrix for filtered SNPs. b Fixation index analysis conducted on japonica (JAP), indica (IND), and mixed populations. c Principal component analysis based on the filtered SNPs, which can unambiguously distinguish japonica from indica accessions, suggesting two distinct genetic backgrounds.

Figure 2

Manhattan plots of SNP markers associated with seed germination index conducted using a fastLMM or b MLMM models. The QTLs are assigned when the peak SNP correlation exceeds the indicator line calculated by Genetic type 1 Error Calculator (GEC) software. Information on QTLs is given in Table 2.

Figure 3

Heatmap of the same ten candidate genes from two different transcriptomic datasets: a GSE115371 (only aerobic lines), b SRP277875 (only Japonica lines). All genes are differentially expressed compared to 0 h in both datasets. Five genes share either same expression patterns in both datasets, i.e., *OsNFYA5*, *PPT1*, while others, i.e., *LOC_Os09g12650*, *YAB1*, exhibit different expression patterns (starred).

Figure 4

Candidate SNPs associated with seed dormancy in Os9BGluc33 (LOC_Os09g33710).