Dynamic Genome-Wide Association Analysis and Identification of Candidate Genes Involved in Anaerobic Germination Tolerance in Rice

Ling Su
South China Agricultural University
Jing Yang
South China Agricultural University
Dandan Li
South China Agricultural University
Ziai Peng
South China Agricultural University
Aoyun Xia
South China Agricultural University
Meng Yang
South China Agricultural University
Lixing Luo
South China Agricultural University
Cuihong Huang
South China Agricultural University
Jiafeng Wang
South China Agricultural University
Hui Wang
South China Agricultural University
Zhiqiang Chen
South China Agricultural University
Tao Guo (guo.tao@vip.163.com)
National Engineering Research Center of Plant Space Breeding, South China Agricultural University, Guangzhou, 510642, China
https://orcid.org/0000-0002-2707-1760

Original article

Keywords: Rice, Anaerobic germination tolerance, Coleoptile, Dynamic GWAS, RNA-seq, Candidate gene

Posted Date: June 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-33743/v1

Version of Record: A version of this preprint was published on January 6th, 2021. See the published version at https://doi.org/10.1186/s12284-020-00444-x.

Abstract

Background: In Asian rice production, an increasing number of countries now choose the direct seeding mode because of rising costs, labour shortages and water shortages. The ability of rice seeds to undergo anaerobic germination (AG) plays an important role in the success of direct seeding.

Results: In this study, we used 2,123,725 single nucleotide polymorphism (SNP) markers based on resequencing to conduct a dynamic genome-wide association study (GWAS) of coleoptile length (CL) and coleoptile diameter (CD) in 209 natural rice populations. A total of 26 SNP loci were detected in these two phenotypes, of which 5 overlapped with previously reported loci (S1_ 39674301, S6_ 20797781, S7_ 18722403, S8_ 9946213, S11_ 19165397), and two sites were detected repeatedly at different time points (S3_ 24689629 and S5_ 27918754). We suggest that these 7 loci *(*$-\log_{10}$ (*P*) value > 7.3271*)* are the key sites that affect AG tolerance. To screen the candidate genes more effectively, we sequenced the transcriptome of the flooding-tolerant variety *YZX* in six key stages, including anaerobic (*AN*) and the oxygen conversion point (*AN-A*), and obtained high-quality differential expression profiles. Four reliable candidate genes were identified: *Os01g0911700* (*OsVP1)*, *Os05g0560900* (*OsGA2ox8*), *Os05g0562200* (*OsDi19-1*) and *Os06g0548200*. Then, q*RT-PCR* and LC-MS/MS-based target metabolite detection technology were used to verify the expression levels of these four candidate genes and the contents of the metabolites they regulate, which specifically responded to AG.

Conclusion: *The four novel candidate genes were associated with gibberellin (GA) and abscisic acid (ABA) regulation and cell wall metabolism under* oxygen-deficiency *conditions and promoted coleoptile elongation while avoiding adverse effects, allowing the coleoptile to obtain oxygen, escape the low-oxygen environment and germinate rapidly. The results of this study improve our understanding of the genetic basis of AG in rice seeds, which is conducive to the selection of flooding-tolerant varieties suitable for direct seeding.*

Background

Cultivated rice (*Oryza sativa* L.) is a staple food for over half of the world's population (Zhang et al. 2019). The world's rice demand is projected to increase by 25% from 2001 to 2025 with a targeted production of 732.5 MT, which is achievable by the addition of 5.9 MT every year. However, based on population growth, the Food and Agricultural Organization estimates that by 2050, the world rice requirement will be 524 MT with an annual increase of 2 MT from the present level of production (Jeon, 2011).

Worldwide, irrigated rice ecosystems comprise 55% of the world's rice-growing area and provide 75% of global rice production (Mahender et al. 2015). However, the current evidence shows a decline in grain yield productivity due to looming threats to natural resources, falling water tables, mounting labour shortages, energy scarcity, increasing input prices and changing climatic conditions (Singh et al. 2013). However, the water-, energy-, and labour-intensive system of transplanted puddled rice is steadily being replaced by direct-seeded rice (DSR) due to the progressive scarcity of these resources (Miro et al. 2017).

DSR can be classified as wet DSR, dry DSR, or water DSR (Mahender et al. 2015). Compared with wet and dry DSR, water DSR (whereby seeds are broadcast into standing water) is more advantageous because it is less labour and time intensive and restrains weed growth (Yamauchi et al. 1993). However, rice is extremely sensitive to anoxia during germination and early seedling growth (Yamauchi et al. 1993; Ismail et al. 2009; Yang et al. 2019). In the case of heavy rainfall during the germination of direct-seeded rice, standing water on the soil surface will significantly affect the seedling rate and thus affect yield (Magneschi et al. 2009). The rice coleoptile is one of the few plant tissues that can grow under hypoxic conditions (Pearce and Jackson, 1991). Anoxia promotes coleoptile elongation and inhibits the growth of young stems and roots (Fred et al. 1981). Under anaerobic conditions, the coleoptile elongates rapidly, breaks through the water surface and transports oxygen to the underwater tissues (Alpi & Beevers, 1983). The coleoptile elongates rapidly to reach the aerobic environment, providing nutrients for the primary leaves and roots and then promoting seedling morphogenesis. Therefore, it is of great significance to explore the mechanism of rice anaerobic germination (AG) tolerance and the morphological and physiological responses of the coleoptile in anaerobic environments to promote the development of rice direct seeding technology.

When rice is in an anaerobic environment, its coleoptile elongation varies widely among different varieties as a typical quantitative characteristic (Perata & Alpi, 1983; Magneschi et al. 2009; Magneschi & Perata, 2009). To determine the genetic basis of this phenomenon, in recent years, quantitative trait loci (QTLs) have been used to identify different populations, and some QTLs related to coleoptile elongation have been found. Angaji (2008) reported four putative QTLs on chromosomes 1, 2, 11, and 12 in a BC_2F_2 population derived from the cross of KHAIYAN and IR64. These QTLs explained 51.4% of the phenotypic variance. Similarly, Khao Hlan On, a anoxia-tolerant accession, was crossed with IR64 to construct BC_2F_2 lines, and five QTLs were found on chromosomes 1, 3, 7, and 9 (Angaji et al. 2009). Among them, *TPP7* (Kretzschmar et al. 2015) was found in *qAG 9 − 2* (Angaji et al. 2010) and shown to participate in trehalose-6-phosphate metabolism, enhance starch mobilization, and improve AG tolerance. Offspring with the tolerance allele showed a high

germination rate, as their seedlings could reach the water surface (Angaji et al. 2010). Septiningsih et al. (2013) used 118 simple sequence repeat (SSR) molecular markers and an $F_{2:3}$ population constructed from IR42 and Ma-Zhan Red to detect 6 QTLs related to seed germination and flooding tolerance under 10 cm flooding stress, and the BC_2F_3 generation was used to verify the major effect locus *qAG7.1* on chromosome 7 and estimate its contribution rate at 31.7%. Lee et al. (2017) used backcross inbred lines (BILs) of Kasalath and Nipponbare to evaluate deep sowing (5 cm sowing depth) performance. With coleoptile length (CL) as an indicator, a QTL location assay identified sites that affected coleoptile elongation on chromosomes 3 and 5, and these sites explained phenotypic variations of 11.8% and 12.0%, respectively (Lee et al. 2017). Yang et al. (2019) used linkage analysis of 192 recombinant inbred lines (RILs) produced from YZX and 02428 to detect 13 stable QTLs on chromosomes 1, 2, 3, 4, 6, 7, 9, 10 and 12, taking the CL, coleoptile diameter (CD), coleoptile surface area (CSA) and coleoptile volume (CV) at six days after submergence and the germination of the seeds of the two generations as indicators.

Most of the existing reports are based on the genetic site mining of populations derived from different parents (such as RILs; NILs; and BILs), but these studies are limited by the genetic information possessed by the parents and by molecular marker density, so their identification of flooding-resistant germination loci related to coleoptile elongation is limited.

Genome-wide association studies (GWAS) based on SSR (Li et al. 2012) or single nucleotide polymorphism (SNP) markers (Huang et al. 2010; Zhao et al. 2011; Huang et al. 2012) have been widely used in model plants, including rice. By identifying high-density SNPs in different germplasm accessions based on second-generation genome sequencing or SNP array approaches, we can obtain a very high resolution (Chen et al. 2014; Yu et al. 2014; Wu et al. 2015), which provides a strong advantage for the mining of complex QTLs. Hsu & Tung (2015) analysed the association of 153 indica, japonica and Australian rice populations by using genome-wide association technology and found that 88 significant SNPs were correlated with the anaerobic response index based on differences in CL between control and submerged rice plants. Among them, only one unique QTL on the long arm of chromosome 1 showed a strong signal in a RIL population from Nipponbare/IR64 hybridization, and this QTL explained approximately 27% of the phenotypic variation (Hsu & Tung, 2015). The study revealed that the *HXK6* gene was located in this genomic region. HXK6 is a double-target mitochondrial and nuclear protein (Cho et al. 2009; Huang et al. 2009). As a glucose sensor (Cho et al. 2009;Granot et al. 2013; Granot et al. 2014), it may be involved in hexose phosphorylation to regulate the expression of important hypoxia-inducible genes, such as *CIPK15* and alcohol dehydrogenase (Yim et al. 2012; Lim et al. 2013; Shingaki - Wells et al. 2014), and therefore improve the survival rate of rice seedlings during submerged germination.

Zhang et al. (2017) carried out GWAS analysis using a mixed linear model with 432 *indica* rice varieties in normal and flooded environments and detected 11 and 9 significant SNPs for flooded coleoptile length (FCL) and flood tolerance index (FTI). Haplotype analysis found that LOC_Os06g03520 contained a protein DUF domain similar to that in genes involved in the energy metabolism pathway (*OsTPP7* and *OsCIPK15*) and was a highly anoxia-induced candidate gene. Nghi et al. (2019) conducted a GWAS on 273 *japonica* rice materials to find new chromosomal regions related to the ability of rice coleoptiles to elongate under flooding conditions. Eleven significant SNPs were found, and 21 candidate genes related to CL were identified.

Coleoptile elongation is the only organic response of rice to anaerobic stress and a reliable characteristic used to study the flooding tolerance of rice genotypes (Setter, 1994; Kato-Noguchi & Morokuma, 2007; Zhang et al. 2017). The ongoing improvements of genome sequencing technology will significantly improve the efficiency of GWAS technology in genetic locus identification by allowing researchers to obtain high-density SNPs by sequencing. In this study, phenotypic data (CL and CD) and 2,123,725 SNP markers were used to perform a dynamic GWAS. The purpose of this study was to detect genetic loci significantly associated with AG in rice, to explore favourable SNP alleles that may be used to breed rice suitable for direct seeding and to find reliable candidate genes to lay a foundation for future study of the molecular mechanism of AG in rice.

Results

Statistical analysis of phenotypic variation in anaerobic germination tolerance

A total of 209 *indica* and *japonica* rice accessions from all over the world (Table S1) were tested for anaerobic germination. CL, CD, CSA and CV on the second, third and fourth days of germination were used as the indexes of submergence tolerance. Abundant genetic variation was detected in this *indica* and *japonica* rice accession population (Fig. 1). Analysis of variance showed that the 209 rice accessions had significant differences in submerged coleoptile phenotype at different time points. At the same time, both genotype and length of submergence time had significant influences on coleoptile phenotype (Table 1). The variation coefficients of CL, CSA and CV were greater than 10% on the 2nd, 3rd and 4th days of submergence, indicating that the CL, CSA and CV of the different varieties were

significantly different. The variation coefficients of CL, CD, CSA and CV were the greatest on the second day of submerging. With increasing submerging time, the variation coefficients of each phenotype decreased gradually, indicating that the differences in each phenotype were significant on the second day of submerging, and the coleoptiles grew rapidly to avoid the anaerobic environment.

Table 1 Phenotypic variation of rice coleoptile during submergence germination.

Phenotype	Environments	Mean ± SE	Range	Coefficient of variation (%)	G × E
CL	AN2d	0.2226 ± 0.0132 cm	0.0062 ~ 0.6463 cm	18.7571	***
	AN3d	0.6629 ± 0.0258 cm	0.0334 ~ 1.7719 cm	13.0543	
	AN4d	1.2068 ± 0.0498 cm	0.1659 ~ 3.8798 cm	12.9767	
CSA	AN2d	0.0266 ± 0.0051 cm^2	0.0045 ~ 0.6463 cm^2	56.1438	
	AN3d	0.1104 ± 0.0048 cm^2	0.0098 ~ 0.3259 cm^2	14.3195	
	AN4d	0.2030 ± 0.0089 cm^2	0.0201 ~ 0.7689 cm^2	13.7986	
CD	AN2d	0.4543 ± 0.0272 mm	0.1092 ~ 0.6597 mm	22.4859	
	AN3d	0.5155 ± 0.0137 mm	0.3908 ~ 0.6837 mm	7.9094	
	AN4d	0.5496 ± 0.0065 mm	0.4452 ~ 1.0680 mm	3.5583	
CV	AN2d	0.0008 ± 0.0001 cm^3	0.0001 ~ 0.0625 cm^3	35.6686	
	AN3d	0.0015 ± 0.0001 cm^3	0.0049 ~ 0.0015 cm^3	18.3201	
	AN4d	0.0030 ± 0.0010 cm^3	0.0003 ~ 0.0419 cm^3	20.8486	

It can be seen from the histogram of frequency distribution (Fig. 2) that after 4 days of continuous oxygen-deficiency treatment, the AG tolerance indexes were consistent with a normal distribution (except CV at 2 and 4 days of submergence), indicating that these traits are regulated by many genes with small effects.

Correlation Analysis Of Phenotypic Traits

Correlation analysis showed correlations among the coleoptile phenotypes after 2 days, 3 days and 4 days of submergence stress (Table 2). Among them, CV-AN2d/CL-AN2d and CSA-AN2d/CSA-AN3d/CSA-AN4d/CD-AN2d/CD-AN3d were correlated, but there was no correlation between CV-AN4d and CD-AN2d. CV-AN2d was negatively correlated with CL-AN3d, CL-AN4d, CD-AN4d, CV-AN3d and CV-AN4d. CL-AN3d/CSA-AN3d/CV-AN3d, CL-AN4d/CSA-AN4d, and CSA-AN3d/CV-AN3d had very significant positive correlations with a P-value > 0.800.

Table 2 Correlation analysis of coleoptile phenotpye.

Trait	CL - AN2d	CL - AN3d	CL - AN4d	CSA - AN2d	CSA - AN3d	CSA - AN4d	CD - AN2d	CD - AN3d	CD - AN4d	CV - AN2d	CV - AN3d	CV- AN4d
CL - AN2d	1											
CL - AN3d	0.521**	1										
CL - AN4d	0.475**	0.739**	1									
CSA - AN2d	0.767**	0.442**	0.405**	1								
CSA - AN3d	0.567**	0.948**	0.725**	0.385**	1							
CSA - AN4d	0.445**	0.657**	0.912**	0.403**	0.668**	1						
CD - AN2d	0.290**	0.268**	0.291**	0.304**	0.353**	0.250**	1					
CD - AN3d	0.215**	0.180**	0.189**	0.266**	0.313**	0.293**	0.632**	1				
CD - AN4d	0.226**	0.148*	0.181**	0.435**	0.195**	0.272**	0.404**	.729**	1			
CV - AN2d	0.066	-0.034	-0.023	0.128	0.042	0.002	0.054	0.005	-0.01	1		
CV - AN3d	0.549**	0.885**	0.710**	0.463**	0.924**	0.687**	0.394**	0.452**	0.354**	-0.032	1	
CV- AN4d	0.217**	0.296**	0.436**	0.205**	0.324**	0.730**	0.01	0.204**	0.175*	-0.016	0.409**	1
** At 0.01 level (2-tailed), the correlation was significant.												
* At 0.05 level (2-tailed), the correlation was significant.												

The dynamic test results of coleoptile phenotypes showed significant correlations among CL and CD measurements at various time points, and these traits play key roles in seed tolerance of anaerobic conditions. This conclusion was also supported by previous work. Therefore, GWAS analysis was performed with CL and CD as phenotypes.

Whole-genome Resequencing And Polymorphism Identification

In this study, the genomes of 209 rice germplasm accessions were resequenced. Burrows-Wheeler Aligner (BWA) software was used to compare the sequencing data with the Japanese Haruka reference genome (IRGSP 1.0.27). Finally, 1,243.37 Gbp of clean data was obtained, and the Q30 reached 93.15%. The average contrast ratio between each sample and the reference genome was 93.06%, the average sequencing depth was 12 ×, and the average coverage of the reference genome was 90.90% (at least one base coverage).

A total of 2,123,725 SNP sites were detected in this dataset. Among them, the SNP types were mainly C:G > T:A and T:A > C:G (Fig. S1a). Using SnpEff software to annotate the detected SNPs, we found 663,818 SNPs with synonymous mutations and 712374 with nonsynonymous mutations (Fig. S1b).

Population Genetics And Evolutionary Analysis

Phylogenetic and population structure analysis

From the phylogenetic tree (Fig. 3a), it can be seen that the population structure used in this experiment is uniform, without strong population stratification. Based on the SNPs, Admixture (Alexander et al. 2009) software was used to analyse the group structure of the research materials. Cross-validation error analysis showed that the error peak was lowest at K = 5, indicating the optimal grouping (Fig. 3c).

The population structure analysis showed no obvious pedigree differentiation in the selected plant materials, confirming that they were suitable for subsequent GWAS analysis. The phylogenetic tree results showed that the selected population could be divided into 5 subgroups, which verified the conclusion that K = 5 was the optimal result in the population structure analysis. The Q matrix with K = 5 was selected for the subsequent association analysis (Fig. 3d).

Based on the SNP data, EIGENSOFT software was used to perform principal component analysis (PCA) (Price et al. 2006) to cluster the samples (Fig. 3b). The results of PCA supported the evolutionary analysis, further confirming that the degree of discreteness of individual kinship in the population was small.

Dynamic genome-wide association analysis of coleoptile phenotypic traits under submergence

Based on the developed high-density SNP marker data, we performed a GWAS on the dynamic changes in two phenotypic traits, CL and CD, to mine new genetic loci associated with AG. The Manhattan plots and quantile-quantile (QQ) plots are shown in Fig. 4 and Fig. 5. Taking $-\log_{10}(P)$ value > 7.3271 as the threshold, 26 significant SNPs were detected (Table 3). Among them, 24 SNPs were significantly associated with CD (9 at AN2d, 1 at AN3d, and 17 at AN4d) and were distributed on chromosomes 1, 2, 3, 5, 6, 7, 8, and 11. Two SNPs were significantly correlated with CL (0 at AN2d, 1 at AN3d, and 2 at AN4d), and they were distributed on chromosomes 1 and 3.

Table 3 Genome-wide significant associations for coleoptiles phenotype under anaerobic germination (AG) using MLM.

Traits	SNP[a]	Chr.	Peak position[b]	*P* value	−log10 (*P*)	Allleles	MAF	Known loci
CD-AN2d	**chr1_39674301**	1	39674301	1.48E-10	9.83	A/C	0.2	Hsu & Tung, 2015; *qCD-1*(ES)(Yang et al. 2019)
	chr2_35360509	2	35360509	7.98E-11	10.1	A/G	0.2	
	chr2_35376690	2	35376690	4.84E-09	8.32	C/A	0.2	
	chr5_7135836	5	7135836	2.32E-11	10.63	C/T	0.1	
	chr5_22190175	5	22190175	2.69E-11	10.57	G/A	0.2	
	chr6_20797781	6	20797781	8.49E-12	11.07	G/A	0.1	Hsu & Tung, 2015
	chr8_13123867	8	13123867	2.65E-11	10.58	G/T	0.1	
	chr8_11595732	8	11595732	5.64E-10	9.25	A/T	0.1	
CD-AN3d	**chr5_27918754**	5	27918754	1.69E-09	8.77	A/G	0.4	Hsu & Tung, 2015
CD-AN4d	chr1_32866906	1	32866906	1.24E-09	8.91	G/A	0.2	
	chr3_27854371	3	27854371	5.11E-16	15.29	A/G	0.3	
	chr3_2828271	3	2828271	4.16E-11	10.38	A/G	0.2	
	chr3_28290634	3	28290634	3.02E-10	9.52	G/A	0.1	
	chr3_28191513	3	28191513	3.83E-10	9.42	C/T	0.1	
	chr3_28336693	3	28336693	3.90E-10	9.41	C/T	0.2	
	chr3_32459722	3	32459722	1.20E-09	8.92	C/T	0.5	
	chr3_7243651	3	7243651	4.89E-09	8.31	C/A	0.2	
	chr3_15571447	3	15571447	2.23E-08	7.65	G/T	0.3	
	chr5_27918754	5	27918754	1.72E-11	10.76	A/G	0.4	Hsu & Tung, 2015
	chr5_18621890	5	18621890	9.76E-11	10.01	A/C	0.1	
	chr5_21585755	5	21585755	1.21E-08	7.92	T/A	0.2	
	chr7_18722403	7	18722403	3.45E-09	8.46	T/C	0.2	*qAG 7.2* (Septiningsih et al. 2013; Hsu & Tung, 2015)
	chr8_22721012	8	22721012	1.31E-09	8.88	A/G	0.2	
	chr8_9946213	8	9946213	2.51E-08	7.6	T/A	0.2	Hsu & Tung, 2015
	chr11_19165397	11	19165397	4.62E-09	8.34	C/A	0.3	*qAG 11*(Jiang et al. 2006; Angaji et al. 2009; Zhang et al. 2017)

CL-AN3d	chr3_24689629	3	24689629	4.71E-08	7.33	T/C	0.4
CL-AN4d	chr1_25099585	1	25099585	1.73E-08	7.76	G/A	0.2
	chr3_24689629	3	24689629	1.83E-08	7.74	T/C	0.4
Note: MAF minor allele frequency;							
a The SNP positions were based on the annotation data on Os-Nipponbare-Reference-IRGSP-1.0 (RAP-DB, http://rapdb .dna.affrc .go.jp/);							
b Position of the SNP showing the most significant association for AG.							

To verify the accuracy of our results, we compared them with previously reported QTLs (Jiang et al. 2006; Yang et al. 2019) related to the control of coleoptile elongation and the survival rate of plants under flooding conditions (Angaji et al. 2010; Baltazar et al. 2014; Septiningsih et al. 2013). We found that some of our QTLs has been identified before in rice submerged germination tolerance. We found that 6 genomic regions were colocalized (grey and orange blocks in Fig. 6). A SNP on chromosome 1 (S1_ 39674301), which was significantly associated with CD in this study, was located in the *qCD-1*(ES) gene interval reported by Yang et al. (2019) as closely related to CD during the early cropping season. One SNP (S6_20797781) located on chromosome 6 was very close to a gene locus detected in another association study (Hsu & Tung, 2015). The physical distance between these two SNPs was only approximately 4 kb. Another SNP (S7_18722403) significantly correlated with CD at AN4d was found inside the genomic interval of *qAG7.2* on chromosome 7; this locus was previously detected using linkage mapping and association mapping (Septiningsih et al. 2013; Hsu & Tung, 2015), validating its genetic effect. On chromosome 8, a SNP (S8_ 9946213) significantly associated with CD was very close to a site detected in another related study as closely related to the anaerobic response index (treated CL vs. control CL; Hsu & Tung, 2015), and the physical distance between them was approximately 95 kb. On chromosome 11, another SNP (S11_ 19165397) significantly associated with CD was located in the genomic region of *qAG 11*, which was closely related to the submerged CL (Hsu and Tung, 2015) and the anaerobic response index (*qAG 11*, Jiang et al. 2006; Angaji et al. 2009).

In addition, some loci were identified at multiple anoxia time points. CD was correlated with SNP S5_ 27918754 at AN3d and AN4d (Fig. 5), as reported in a previous study (Hsu & Tung, 2015). In addition, CL was corelated with SNP S3_24689629 (Fig. 4) at AN3d and AN4d, and this result had not been reported previously.

Identification Of Candidate Genes

We selected 7 loci from the GWAS results, including S3_ 24689629 and S5_ 27918754, which were repeatedly detected for CD and CL, respectively, at both AN3d and AN4d, indicating that these two loci were highly correlated with coleoptile elongation and thickening under hypoxic conditions. In addition, S1_39674301, S6_20797781, S7_18722403, S8_9946213, and S11_19165397 (Hsu & Tung, 2015; Yang et al. 2019; Septiningsih et al. 2013; Zhang et al. 2017; Jiang et al. 2006; Angaji et al. 2009) were consistent with previous research results and were included. Based on the linkage disequilibrium (LD) decay distance (100–350 kb) (Fig. S2), we detected a total of 106 differentially expressed genes (DEGs) in these seven trait-associated SNP sites. To more effectively identify candidate genes related to AG in rice seeds, RNA-seq analysis was carried out on RNA extracted from germinated seeds of the flooding-tolerant variety YZX at AN0d, AN2d, AN3d, AN4d, AN3dA1d, and A3dAN1d. We normalized the FPKM values of these DEGs as follows (Fig. 6): $\log_2$ FC (AN3dA1d/AN3d), $\log_2$ FC (AN3dA1d/AN4d), $\log_2$ FC (A3dAN1d/A3d), and $\log_2$ FC (A3dAN1d/A4d). Based on the candidate gene selection criteria $|\log_2 FC| \geq 1$, FDR ≤ 0.05 and FPKM ≥ 1 (special: log2 FC (AN3dA1d/AN3d), $\log_2$ FC (AN3dA1d/AN4d) ≤ 1; $\log_2$ FC (A3dAN1d/A3d), and $\log_2$ FC (A3dAN1d/A4d) ≥ 1), a total of 4 DEGs were obtained.

According to the gene annotation information of the rice reference genome, we concluded that these four DEGs, gibberellin 2-beta-dioxygenase (*OsGA2ox8*, *Os05g0560900*), berberine bridge enzyme-like 13 (*Os06g0548200*), protein dehydration-induced 19 (*OsDi19-1*, *Os05g0562200*), and B3 domain-containing protein VP1 (*OSVP1*, *Os01g0911700*), had major impacts on AG in rice seeds. The results of qRT-PCR (Fig. 7) were similar to the results of RNA-sEq. These four genes were specifically expressed under hypoxic conditions, their expression levels increased with increasing hypoxia time, and their expression levels decreased after transfer to aerobic conditions. These results indicated that these four DEGs had strong correlations with submerged germination in rice.

Four New Genes Related To The Anaerobic Germination Phenotype

To further verify the associations between the candidate genes and the germinating coleoptile phenotype under AG conditions, we performed a full-length sequencing analysis of the four candidate genes in materials with extreme phenotypes (11 high-AG materials and 11 low-AG materials) (Fig. 8).

Os01g0911700 (*OsVP1*) encodes the VP1 protein, which contains the B3 domain and is a key transcription factor in the abscisic acid (ABA) signal transduction pathway. Among the 11 high-AG materials and 11 low-AG materials, we identified 11 allelic variants and 3 haplotypes of this gene (Fig. 8a). Of the 11 allelic variations, 10 included mutation sites within the gene, and these sites were mainly distributed in the first, third and sixth exon regions. The 11 high-AG varieties were different from the low-AG varieties at these 10 loci. The haplotype analysis of the extreme materials showed that Hap.2 (compared to Hap.3, all the variable sites except + 471, downstream of the start codon, were different) was associated with the high-AG phenotype based on the traits of the coleoptile at AN4d. The average CL value was 2.932 cm. Plants carrying Hap.3 (in which none of the 11 variable sites were mutated) all had the low-AG phenotype, and the average CL value at AN4d was 0.257 cm (Table S4). Therefore, we think that the mutations at the remaining 10 sites have a greater impact on the AG phenotype than the mutation at site + 471 (downstream of the start codon).

The *Os05g0560900* (*OsGA2ox8*) gene encodes a gibberellin 2-beta-dioxygenase, which is involved in GA catabolism. From the gene sequences of the 11 high-AG materials and 11 low-AG materials, we identified 17 allelic variations (Fig. 8b), which were mainly located in the first intron region. Two were located in the first and third exon regions, and the first was located at position 431 (C → G) after the start codon, causing a missense mutation of glycine (G) to cysteine; the second was located at position 1323 (A → T) after the start codon, causing a missense mutation from glutamine (Q) to leucine (L). Haplotype analysis showed 8 haplotypes in these 22 extreme materials (Table 4). Among them, Hap.2 was associated with the high-AG phenotype. From the perspective of gene sequence, the high-AG phenotype appeared to be caused by a heterozygous mutation or deletion at + 1235 (downstream of the start codon) and was apparent in seedlings at AN4d, when the average CL was 2.931 cm. Hap.5 and Hap.6 had missense mutations in the first exon (+ 431) and the third exon (+ 485); plants with these haplotypes showed both high-AG and low-AG phenotypes, and their average CL values at AN4d were 2.412 and 2.274 cm, respectively. In contrast, Hap.8 was associated with the low-AG phenotype, and the average CL value of plants with this haplotype at AN4d was 0.257 cm (Table S4).

Os05g0562200 (*OsDi19-1*) encodes the dehydration-induced Dil9 protein, which contains two zf-Di19 and one Di19_C domains. From the gene sequences of 22 (11 high-AG and 11 low-AG) materials, we identified 15 allelic variants, all of which were located within the gene (Fig. 8c). The gene sequences of the low-AG material were mainly similar to the sequence of the reference genome (IRGSP-1.0.27), and the gene sequences of the high-AG material harboured different mutations at these 15 loci. We speculated that these 15 allelic variations were closely related to the high-AG phenotype. In particular, three missense mutations were located in the UTR, and two were located in the third and fourth exon regions. The mutation at 120 bases upstream of the start codon is T → G, the mutation at 80 bases upstream of the start codon is C → T, the mutation 1395 bases downstream of the start codon is A → G, the mutation at1860 bases downstream of the start codon is T → C (isoleucine to threonine), and the mutation at 2727 bases downstream of the start codon is G → A. Haplotype analysis showed 7 haplotypes among these 22 extreme materials (Table 4). Hap.2 was the reference genome sequence, and Hap.2 plants mainly showed the low-AG phenotype, with an average CL value at AN4d of 0.288 cm. The average CL values were 2.391 and 2.269 cm, respectively (Table S4).

Os06g0548200 encodes berberine bridge enzyme (BBE). We selected 11 high-AG and 11 low-AG materials and identified 9 allelic variations in their gene sequences (Fig. 8d) that were potentially related to the AG phenotype. One G → A mutation was located 31 bases upstream of the start codon; a missense mutation was located 645 bases downstream of the start codon (C → G, proline (P) to arginine (R)); a missense mutation (T → G) located 1607 bases downstream of the start codon resulted in an aa substitution from isoleucine (I) to serine (S); a missense mutation (A → T) was located 1685 bases downstream of the start codon (asparagine (N) to I). These four allelic variants had only one genotype (mutant type) in the high-AG materials, while there were two different genotypes (mutant and reference genotypes) in the low-AG materials, and these had polymorphisms located in the UTR and outside the exon regions. Haplotype analysis showed 10 haplotypes in these 22 extreme materials (Table 4), of which Hap.5 showed a high-AG phenotype, and its average CL at AN4d was 2.274 cm. Hap.8 showed a low-AG phenotype, and its average CL at AN4d was 0.273 cm (Table S4). Hap.5 had mutations at -31, + 645, +1607, and + 1685, and Hap. 8 had the opposite pattern, so we believe that these four sites may be closely related to the high-AG phenotype.

Metabolic Spectrum Detection Of Rice Seeds Under Anaerobic Germination

LC-MS/MS was used to obtain the mass spectrum data of samples, and qualitative and quantitative analysis was carried out on the basis of a self-established database (MWDB) by Metware Company and public metabolite information database. In total, 730 metabolites were identified, including 32 substances and their derivatives (Table S2).

Through the detection of metabolites in the seeds of the YZX rice variety during AG, we found that the contents of metabolites regulated by the four genes identified above were specifically correlated with hypoxic stress. Among them, we detected three biologically active GAs, GA_9, GA_{15}, and GA_{20}, which are related to the regulation of GA catabolism by the *OsGA2ox8* (*Os05g0560900*) gene. The contents of GA_9 and GA_{15} increased within 2 days after submergence, and the content at AN4d decreased compared to that at AN3d; the content of GA_{20} decreased to its lowest value within 2 days after flooding and then did not change (Table S3, Fig. 10). *OsVP1* (*Os01g0911700*) and *OsDi19-1* (*Os05g0562200*) participated in the regulation of ABA-related phenotypes. Quantitative results showed that ABA content decreased within 3 days after flooding and then suddenly increased on the fourth day (Table S3, Fig. 10). The BBE-like enzyme encoded by *Os06g0548200* is involved in plant phenylpropanoid metabolism. Quantitative results showed that the contents of shikimic acid and Phe increased with increasing submerging time (Table S3, Fig. 9).

Discussion

In rice production in Asia, due to rising costs, labour shortages and water shortages, farmers often choose to plant directly in rain-fed and irrigation systems (Miro & Ismail, 2013). In this case, the ability of rice to withstand flooding during germination is crucial. Rice seeds can germinate under hypoxic or even anaerobic conditions and extend the coleoptile to some extent (Biswas & Yamauchi, 1997; Ella & Setter, 1999), but growth during AG and the early seedling stage is extremely sensitive to anaerobic conditions (Yamauchi et al. 1993; Ismailet al. 2009; Angaji et al. 2010). Only tolerant genotypes have the ability to rapidly extend the coleoptile and form roots under submerged conditions in the field (Ismail et al. 2009). In contrast, the coleoptiles of sensitive genotypes grow more slowly, and the seedlings cannot develop further. It has been speculated that most rice genotypes can initiate germination but cannot complete it. Similar studies have reported the use of GWAS methods to mine functional genes for AG in rice, but few accessions have been used, and dynamic GWAS analysis examining the germination stage and evidence for targeted genes have been lacking (Hsu & Tung, 2015; Zhang et al. 2017; Nghi et al. 2019). In this study, a high-density SNP marker set based on resequencing was used to perform a GWAS on the CL and CD of 209 diverse rice materials with RNA-Seq based on a differential expression profiling strategy. We identified 4 notable genes, and their expression levels were consistent with changes in the coleoptile phenotype under AG conditions. Furthermore, we used LC-MS/MS extensive targeted metabolic detection technology to verify the contents of metabolites related to the regulatory pathways associated with these four genes to support their importance in rice AG tolerance. These results may further enhance our understanding of AG in rice and help to explore the genetic mechanisms of rice AG during flooding.

The plant genome contains a large number of genes encoding BBE, and the BBE protein family is widespread in the plant kingdom and involved in plant phenylpropanoid metabolism. Phenylpropanoid metabolism is an important pathway for the synthesis of secondary metabolites in plants. All substances containing a phenylpropane skeleton are directly or indirectly produced by this pathway. These compounds have multiple functions in plants (Schadel et al. 1981). In recent years, in the process of genome sequencing, many genes encoding BBE-like enzymes have been identified in plants and bacteria. In the model plant *Arabidopsis thaliana*, 28 BBE-like genes referred to as *AtBBE-like 1–28* (Wallner et al. 2013) were identified and found to be expressed in certain developmental stages of Arabidopsis, such as elongation and maturation, proliferation and embryonic development (Winter et al. 2007). Daniel et al. (2015) found that the two enzymes At-BBE-like 13 and At-BBE-like 15 may be involved in the mobilization and oxidation of the monolignols required for the polymerization process of the plant cell wall (such as lignification), while the BBE-like 15 enzyme family and potentially members of other related families work in concert with them to mobilize building blocks from their storage forms, participate in the cell wall formation required by Arabidopsis growth, and respond to stressors. Our dynamic GWAS correlation analysis found that *Os06g0548200* is located in a genomic region significantly correlated with CD and is highly expressed in the coleoptile under hypoxic conditions. In the early stage, we found that fructose hexaphosphate (F-6-P) is an important metabolite produced in response to AG in rice seeds. F-6-P provides energy through glycolysis and participates in the production of compounds, such as phosphoenolpyruvate (PEP), through the TCA cycle. PEP then enters phenylpropanoid metabolism through the synthesis of shikimic acid and phenylalanine (Phe) (Fig. 9). The results of our qualitative and quantitative metabolite analysis of YZX seeds during AG showed that the contents of shikimic acid and Phe increased with increasing submergence time (Table S3, Fig. 10), consistent with the expression level of *Os06g0548200*.

The *OsGA2ox8* (*Os05g0560900*) gene is located in a region that was significantly associated with CD at both AN3d and AN4d. It encodes gibberellin 2-beta-dioxygenase, which controls participation in gibberellin (GA) catabolism. GAs are a class of endogenous plant hormones that regulate growth and development throughout the life cycle of higher plants (Hooley, 1994), including seed germination, hypocotyls and stem elongation, leaf extension, epidermal trichome development, flowering time, floral organ development, and fruit ripening (Yamaguchi

et al. 1998; Itoh et al. 1999). The concentration of GA is regulated through both synthesis and inactivation; to prevent its gradual accumulation in the plant and ensure the normal operation of the chemical communication system, a decomposition mechanism removes active GA. The main GA inactivation pathway is through GA 2-oxidation, which regulates GA concentration. GAs play important roles in plant growth and development (Rieu et al. 2008). In pea (*Pisum sativum*), loss of function in *GA2ox1* can increase the content of GA_{20} in mature seeds. After germination, this compound is activated by 3β hydroxylation and increases seedling elongation (Ross et al. 1995; Lester et al. 1999; Martin et al. 1999). In addition, loss of function in *GA2ox2* in Arabidopsis increased the level of GA_4 in seeds germinated under dark conditions and promoted the germination of mutants (Yamauchi et al. 2007). Double mutants of *GA2ox7* and *GA2ox8* had higher germination rates and longer hypocotyls in the presence of pyrimidinol, a GA biosynthesis inhibitor (Schomburg et al. 2003) In rice, different members of the *GA2ox* family regulate GA levels during flowering, tillering, and seed germination. This regulation can involve family members or a single gene acting alone (Lo et al. 2008). In the detection of metabolites in YZX seeds under anaerobic germination conditions, we detected three kinds of direct precursors of GA with biological activity: C_{19}-GA(GA_9 ,GA_{20}), C_{20}-GA(GA_{15})(Ross et al. 1995). Their content gradually decreased with the increase of submerging time, while the expression of *OsGA2ox8* (*Os05g0560900*) gene increased with the increase of submerging time. We speculate that *OsGA2ox8* gene may be involved in GA catabolism and inactivate the direct precursors of three kinds of bioactive GA, so as to regulate the endogenous level of bioactive GAs (such as GA_1, GA_4) and the coleoptile elongation of rice seeds.

The plant hormone ABA plays central roles in seed maturation, germination and adaptation to abiotic environmental stress (Leung & Giraudat, 1998). Mohanty et al. (2012) found that ABA is a positive regulator of rice seed germination under flooding and that it seems to interact with ethylene to control coleoptile elongation. *OsVP1* (*Os01g0911700*) encodes the protein VP1, which contains the B3 domain and is a key transcription factor in the ABA signal transduction pathway. In our dynamic GWAS analysis, *Os01g0911700* was found in an area significantly correlated with CD at AN2d. We speculate that *Os01g0911700* may participate in ABA signal transduction during AG (Fig. 9), regulating seed germination in low-oxygen conditions.

Osdi19-1 (*Os05g0562200*) encodes the dehydration-induced Di19 protein, a member of a small family of plant transcription factors, and is reported to be involved in abiotic stress (Liu et al. 2013; Wang et al. 2016). *OsDi19-4*, a member of the rice Di19 protein family, has been reported to act as a transcription factor. Through the *OsCDPK14 – OsDi19-4 – OsASPG1*/*OsNAC18* regulatory pathways, it not only indirectly regulates a large number of plant stress and ABA response genes but also, by interacting with their promoters, directly regulates certain ABA-responsive genes (such as *OsASPG1* and *OsNAC18*), which further modulate ABA-related phenotypes (Wang et al. 2016). In our dynamic GWAS analysis, *OsDi19-1* was detected at both AN3d and AN4d, and its expression level increased with time under anaerobic conditions and when plants were transferred to anaerobic from aerobic conditions. Under aerobic conditions, its expression level decreased; thus, it specifically responded to a hypoxic environment. In our dynamic GWAS analysis, *OsDi19-1* was detected at both AN3d and AN4d, and its expression increased with time under anaerobic conditions and decreased after a change from anaerobic to aerobic conditions, showing a specific response to the anaerobic environment. We quantitatively analysed ABA in YZX seeds at different time points under anaerobic germination conditions and found that the content of ABA gradually decreased in the first 3 days after flooding but suddenly increased on the 4th day. Under timed oxygen-deficiency conditions, *OsDi19-1* directly or indirectly regulates ABA-related response genes, increasing ABA content, which in turn regulates the coleoptile phenotype (Table S3, Fig. 9).

Three of the 4 candidate genes identified by our dynamic GWAS association analysis results (*Os05g0560900*, *Os01g0911700*, *Os05g0562200*) were related to the plant hormones ABA and GA, which indicated that ABA and GA play important roles in coleoptile elongation during AG in rice seeds. Quantitative analysis of GA and ABA in submerged YZX seeds showed that GA content decreased to a plateau over time, and ABA content decreased first and then increased (Table S3, Fig. 10). After 0–3 days of submergence, the GA content was 2–3 times that of ABA; after 4 days of flooding, the GA content reached a minimum, and the ABA content increased to 185 times that of GA. These results indicate that although GA and ABA both play roles in coleoptile elongation under submerged conditions, their regulatory responses have temporal and spatial differences. How these three candidate genes participate in regulating the GA and ABA signalling regulatory pathways will be further explored in future work.

Conclusion

In this study, high-density SNP markers based on resequencing were used to perform dynamic GWAS on the coleoptile phenotypes of 209 natural rice populations, and four reliable candidate genes were identified. These genes were involved in GA and ABA signalling and cell wall metabolic processes, showed differential expression under AG conditions, and promoted coleoptile elongation, allowing seedlings to escape immersion by extending the coleoptile out of the anaerobic environment to obtain oxygen and support rapid germination.

Materials And Methods

Plant material and growth conditions

The mapping population consisted of 209 rice accessions. The majority came from Southern China, whereas others were from Japan, the Philippines, Vietnam, Indonesia, Pakistan, India, the United States and other countries (Table S1). Seeds of the population were planted in a paddy field at South China Agricultural University, Guangzhou (23°16' N, 113°8' E), during the late cropping season (July - November) in 2018. The spaces between rows and between plants were 20 and 20 cm, respectively. Thirty-six plants of each accession were grown in 6 rows with 6 plants per row. Considering that seed maturity affects anaerobic germination (AG), six individual plants in the middle of each block were harvested independently on the 40th day after heading in the late cropping season. Crop management and disease and insect pest control were performed as locally recommended. The harvested seeds were dried in a heated air dryer at 42 °C for 5 days and then stored at − 20 °C (Yang jing et al. 2019).

Phenotyping

From the six independently harvested plants, 30 healthy, fully mature seeds were selected for each accession. The seeds of each experimental cultivar were placed in an oven at 50 °C for 5 days to break dormancy, surface-sterilized with 0.2% $HgCl_2$ for 5 min and then washed thrice with sterile distilled water. Five seeds were placed in one sterile centrifuge tube (50 mL), which was filled with distilled water to create anaerobic conditions. Germination took place at 30 °C in a growth chamber in the dark. After 2, 3, and 4 days, a WinRHIZO (Regent Instruments Inc., Québec, Canada) root image analysis system was used to measure the coleoptile length (CL), coleoptile surface area (CSA), coleoptile volume (CV), and coleoptile diameter (CD). Three independent, replicated experiments were performed for each rice accession. All treatments were performed in parallel. Statistical analysis was performed with SPSS (Statistical Analysis System, version 23.0) and Microsoft Excel.

Whole-genome Resequencing And Variation Detection

The genomes of all 209 accessions were sequenced on the Illumina HiSeq 2500 Sequencing Systems Platform (Illumina Inc. USA), with an average coverage of approximately 12×. Raw reads were processed to obtain an average quality score (QS) per read ≤ 30 by trimming the 3’ ends of the reads using SICKLE (https://github.com/najoshi/sickle). High-quality reads were aligned against the rice reference genome (IRGSP 1.0) (Kawahara et al. 2013) for genotype calling using Burrows - Wheeler Aligner (BWA) (version 0.7.5a).

Single nucleotide polymorphisms (SNPs) were mainly detected by the GATK (McKenna et al. 2010) software kit. According to the positioning results of clean reads in the reference genome, Picard (http://sourceforge.net/projects/picard/) was used to filter the redundant reads (mark duplicates) to ensure the accuracy of the detection results. Finally, low-quality markers, deletion markers < 25% and minor allele frequency (MAF) > 5% were filtered, and the final set of mutation loci was obtained.

Population Structure Analysis And Genome-wide Association Mapping

Principal component analysis (PCA), construction of a neighbour-joining (NJ) tree, determination of linkage disequilibrium (LD) decay level and kinship analysis among the landraces were performed based on SNPs. The population structure of the 209 varieties was estimated with PCA by using the software EIGENSTRAT (Tamura et al. 2007). PHYLIP version 3.695 software (http://evolution.genetics.washington.edu/phylip/getme-new1.html) was used to construct the NJ tree on the basis of similarity measures, and MEGA V6.0 was used to observe the NJ tree (Purcell et al. 2007). A Q matrix was obtained from the membership probability of each variety using ADMIXTURE version 1.22 software (Hardy et al. 2002). The Q matrix was used for further association mapping. The Loiselle algorithm was chosen to construct a kinship matrix (K) with the software SPAGeDi (Kang et al. 2010). Moreover, all negative kinship values were set to zero.

Based on the developed high-density SNP molecular marker data, this study used TASSEL 4.0 (Bradbury et al. 2007), FaST-LMM (Lippert et al. 2011), and EMMAX (Yang et al. 2014) for correlation analysis. Association analysis was implemented under a mixed linear model (MLM), which could be described as $y = X\alpha + S\beta + K\mu + e$. X and y represented the genotype and phenotype, α represented the SNP effects, β represented the effects of population structure, μ was the vector of kinship background effects, e corresponded to residual effects, S was the PCA matrix relating y to β, and K represented the related kinship matrix (Wen et al. 2014; Zhang et al. 2017;Yang et al. 2019).

The whole set of 2,123,725 SNPs was used in association mapping with a MAF criterion of 5%. A genome-wide threshold of -log(*P*) = 7.3271 was calculated from the formula “-log10(0.1/effective number of SNPs)” (Wu et al. 2015).

Ld Structure And Haplotype Analysis

Using the software plink2 (Purcell et al. 2007), on the same chromosome, the LD between pairs of SNPs within a certain distance range (1,000 kb) was calculated. When the LD value of the two markers dropped to half of the maximum value (R^2 = 0.3), the interval was considered a candidate region where potential causal variants might reside. Gramene Mart (http://www.gramene.org/) was used to search for candidate genes in the target region (Hsu & Tung, 2015).

Haplotyping of the identified candidate genes was carried out using DnaSP v5 (Librado and Rozas, 2009). Candidate gene sequence analysis was based on the rice reference genome (IRGSP 1.0) to capture the nucleotide variation in the whole region of each target gene.

Validation Of Gene Expression By Qrt-pcr

For qRT-PCR, total RNA was extracted from rice coleoptiles after 2, 3, and 4 days of germination using a MiniBEST Plant RNA Extraction kit (Takara Bio, Inc., Japan). According to the method of Guo et al. (2019), the candidate genes were analysed by qRT-PCR, and NCBI Primer BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to design gene-specific primers. The primer sequences of the 4 candidate genes are shown in Table S5.

Detection of metabolites in seeds under anaerobic conditions by LC-MS/MS

To understand the metabolic process of rice seed germination under anaerobic conditions, YZX was selected for metabolite detection among the 209 accessions, and LC-MS/MS was used to detect metabolites. The test seeds were placed in an oven at 50 °C for 5 days to break dormancy, surface sterilized with 0.2% $HgCl_2$ for 5 min and washed with sterile distilled water 3 times. Groups of 5 seeds were placed in sterile centrifuge tubes (50 ml), which were filled with distilled water to achieve anaerobic conditions and germinated in a constant-temperature incubator at 30 °C in darkness. After 2, 3 and 4 days of AG, three biological repeats were taken from each treatment, and dry seeds were used as a control. The methods of Chen et al. (2013) were used for sample preparation, extraction and qualitative and quantitative determination of metabolites. The metabolite database used included the MWDB database established by Metware company, the metabolite information public database (MassBank (http://www.massbank.jp), KNApSAcK (http://kanaya.naist.jp/KNApSAcK), Human Metabolome Database (HMDB; http://www.hm db.ca), MoTo DB (http://www.ab.wur.nl/moto), and METLIN (http://metlin.scripps.edu/index.php)).

Abbreviations

AG
anaerobic germination; CL:coleoptile length; CSA:coleoptile surface area; CD:coleoptile diameter; CV:coleoptile volume; DSR:direct-seeded rice; GWAS:genome-wide sequencing analysis; LC-MS/MS:liquid chromatography tandem mass spectrometry.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and material

All data generated or analyzed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Funding

Financial support for this research was provided in part by a grant from the Research and Development Plan for Key Areas in Guangdong province (No. 2018B020206002), the National Key Research and Development Project (No. 2017YFD0100104), and the earmarked fund for China Agriculture Research System (No. CARS-01-17). The funder had no role in the experimental design, data collection and analysis or preparation of the manuscript.

Authors' contributions

TG, ZC, HW and LS designed the project and LS performed all the experiments and wrote the manuscript. JY, DL, ZP, AX, MY, LL, CH and JW assisted in conducting the experiments and analyzing the data. HW and ZC provided the direction for the study and the corrections of the manuscript. All authors read and approved the final manuscript.

Author details

National Engineering Research Center of Plant Space Breeding, South China Agricultural University, Guangzhou 510642, China

Acknowledgments

We wish to thank all the students who participated in this project helping with the seed germination experiment. We thank Guangzhou Genedenovo and Metware Biotechnology Co., Ltd., for assisting in sequencing and bioinformatics analyses. We also thank American Journal Experts (AJE) for its linguistic assistance during the preparation of this manuscript.

References

1. Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664 ()
2. Alpi A, Beevers H (1983) Effects of O_2 concentration on rice seedlings. Plant Physiol 71(1):30–34
3. Angaji SA (2008) Mapping QTLs for submergence tolerance during germination in rice. Afr J Biotechnol 7:2551–2558. doi:10.1007/s00122-013-2057-1
4. Angaji SA, Septiningsih EM, Mackill DJ, Ismail AM (2009) QTLs associated with tolerance of flooding during germination in rice (*Oryza sativa* L.). Euphytica 172:159–168. doi:10.1007/s10681-009-0014-5
5. Angaji SA, Septiningsih EM, Mackill DJ, Ismail AM (2010) QTLs associated with tolerance of flooding during germination in rice (Oryza sativa L.). Euphytica 172:159 –159 68 ()
6. Biswas JK, Yamauchi M (1997) **Mechanism of seedling establishment of direct-seeded rice (***Oryza sativa* **L.) under low land conditions**. BotBullAcadSin 38:29–32 ()
7. 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. 10.1093/bioinformatics/btm308 () :. **doi**
8. Chen HD, Xie WB, He H, Yu HH, Chen W, Li J, Yu R, Yao Y, Zhang WH, He YQ, Tang XY, Zhou FS, Deng XW, Zhang QF (2014) A high-density SNP genotyping array for rice biology and molecular breeding. Mol Plant 7:541–553. doi:10.1093/mp/sst135
9. Chen W, Gong L, Guo Z, Wang W, Zhang H, Liu X, Yu S, Xiong L, Luo J (2013) A novel integrated method for large-scale detection, identification, and quantification of widely targeted metabolites: Application in the Study of Rice metabolomics. Mol Plant 6:1769–1780 ()
10. Cho JI, Ryoo N, Eom JS, Lee DW, Kim HB, Jeong SW, Lee YH, Kwon YK, Cho MH, Bhoo SH, Hahn TR, Park YI, Hwang I, Sheen J, Jeon JS (2009) Role of the Rice Hexokinases OsHXK5 and OsHXK6 as Glucose Sensors. Plant Physiol 149:745–759. doi:10.1104/pp.108.131227
11. *Daniel B*, *Pavkov-Keller T*, *Steiner B*, *Dordic A*, *Gutmann A*, *Nidetzky B*, *Sensen CW*, *van der Graaff E*, *Wallner S*, *Gruber K*, *Macheroux P* (*2015*) *Oxidation of monolignols by members of the berberine bridge enzyme family suggests a role in plant cell wall metabolism. J. Biol. Chem. 290:18770–18781*
12. Ella ES, Setter TL (1999) Importance of seed carbohydrates in rice seedling establishment under anoxia. Acta Hort 504:209–216 ()
13. Fred TT, Chen CC, Garry NM (1981) Morphological development of rice seedling in water at controlled oxygen levels. Agron J 73:556–560 ()

14. Granot D, David-Schwartz R, Kelly G (2013) Hexose kinases and their role in sugar-sensing and plant development. Front Plant Sci doi. 10.3389/fpls.2013.00044
15. Granot D, Kelly G, Stein O, David-Schwartz R (2014) Substantial Roles of Hexokinase and Fructokinase in the Effects of Sugars on Plant Physiology and Development. J Exp Bot 65:809–819. doi:10.1093/jxb/ert400
16. Guo T, Yang J, Li DX, Sun K, Luo LX, Xiao WM, Wang JF, Liu YZ, Wang S, Wang H, Chen ZQ (2019) **Integrating GWAS, QTL, mapping and RNA-seq to identify candidate genes for seed vigor in rice (***Oryza sativa* **L.)**. Mol Breeding 39:87. 10.1007/s11032-019-0993-4 () :. **doi**
17. Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620 ()
18. *Hooley R* (*1994*) *Gibberellins: Perception, Transduction and Responses*. *Plant Mol Biol. 26*:*1529–1555*
19. Hsu SK, Tung CW (2015) Genetic mapping of anaerobic germination associated QTLs controlling coleoptile elongation in rice. Rice 8:38. doi:10.1186/s12284-015-0072-3
20. Huang S, Taylor NL, Narsai R, Eubel H, Whelan J, Millar AH (2009) Experimental Analysis of The Rice Mitochondrial Proteome, Its Biogenesis, and Heterogeneity. Plant Physiol 149:719–734. doi:10.1104/pp.108.131300
21. Huang XH, Wei XH, Sang T, Zhao Q, Feng Q, Zhao Y, Li CY, Zhu CR, Lu TT, Zhang ZW, Li M, Fan DL, Guo YL, Wang AH, Wang L, Deng LW, Li WJ, Lu YQ, Weng QJ, Liu KY, Huang T, Zhou TY, Jing YF, Li W, Lin Z, Buckler ES, Qian Q, Zhang QF, Li JY, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. Nat Genet 42:961–967
22. Huang XH, Zhao Y, Wei XH, Li CY, Wang AH, Zhao Q, Li WJ, Guo YL, Deng LW, Zhu CR, Fan DL, Lu YQ, Weng QJ, Liu KY, Zhou TY, Jing YF, Si LZ, Dong GJ, Huang T, Lu TT, Feng Q, Qian Q, Li JY, Han B (2012a) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. Nat Genet. 2012;44:32–9
23. Ismail AM, Ella ES, Vergara GV, Mackill DJ. (2009) **Mechanisms associated with tolerance to flooding during germination and early seedling growth in rice (***Oryza sativa*). Ann Bot 103:197–209 ()
24. Itoh H, Tanaka-Ueguchi M, Kawaide H, Chen X, Kamiya Y, Matsuoka M (1999) The gene encoding tobacoo gibberellin 3-hyhydroxylase is expressed at the site of GA action during stem elongation and flower development. Plant J 20:15–24 ()
25. Jeon JS, Jung KH, Kim HB, Suh JP, Khush GS (2011) Genetic and molecular insights into the enhancement of rice yield potential. J Plant Biol 54:1–9 ()
26. Jiang L, Liu S, Hou M, Tang J, Chen L, Zhai H, Wan JM (2006) **Analysis of QTLs for seed low temperature germinability and anoxia germinability in rice (***Oryza sativa* **L.)**. Field Crops Res 98:68–75. 10.1016/j.fcr.2005.12.015 () :. **doi**
27. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348 –348 54 ()
28. Kato-Noguchi H, Morokuma M (2007) Ethanolic fermentation and anoxia tolerance in four rice cultivars. J Plant Physiol 164:168–173. doi:10.1016/j.jplph.2005.09.017
29. Kawahara Y, **de la** Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu JZ, Zhou SG, Childs KL, Davidson RM, Lin HN, 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(1):4. **doi**:10.1186/1939-8433-6-4
30. Kretzschmar T, Pelayo MAF, Trijatmiko KR, Gabunada LFM, Alam R, Jimenez R, Mendioro MS, Loedin IHS, Sreenivasulu N, Serres JB, Ismail AM, Mackill DJ, Septiningsih EM (2015) A trehalose-6-phosphate phosphatase enhances anaerobic germination tolerance in rice. Nat Plants 1:15124. doi:10.1038/nplants.2015.124
31. Lee HS, Sasaki K, Kang JW, Sato T, Song WY, Ahn SN (2017) Mesocotyl elongation is essential for seedling emergence under deepseeding condition in rice. Rice 10:32. doi:10.1186/s12284-017-0173-2
32. *Lester DR*, *Ross JJ*, *Smith JJ*, *Elliott RC*, *Reid JB* (*1999*) *Gibberellin 2-oxidation and the* SLN *gene of* Pisum sativum. *Plant J. 19*:*65–73*
33. *Leung J*, *Giraudat J*, (*1998*) *Abscisic acid signal transduction*. *Annu. Rev. Plant Physiol. Plant Mol. Biol.49*:*199–222*
34. Li XB, Yan WG, Agrama H, Jia LM, Jackson A, Moldenhauer K, Yeater K, Anna M, Wu DX (2012) Unraveling the complex trait of harvest index with association mapping in rice (*Oryza sativa* L.). PLoS ONE 7:e29350
35. Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. 10.1093/bioinformatics/btp187 () :. **doi**
36. Lim MN, Lee SE, Yim HK, Kim JH, Yoon IS, Hwang YS (2013) Differential Anoxic Expression of Sugar-Regulated Genes Reveals Diverse Interactions Between Sugar and Anaerobic Signaling Systems in Rice. Mol Cells 36:169–176. doi:10.1007/s10059-013-0152-4

37. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D (2011) Fast linear mixed models for genome-wide association studies. Nat Methods 8:833–835
38. *Liu WX*, *Zhang FC*, *Zhang WZ*, *Song LF*, *Wu WH*, *Chen YF* (*2013*) Arabidopsis *Di19 functions as a transcription factor and modulates PR1, PR2, and PR5 expressions in response to drought stress. Molecular Plant 6:1487–1502*
39. Lo SF, Yang SY, Chen KT, Hsing YI, Zeevaart JA, Chen LJ, Yu SM. (2008) A novel class of gibberellin 2-oxidases control semidwarfism, tillering, and root development in rice. Plant Cell 20:2603–2618. 10.1105/tpc.108.060913 () :. **doi**
40. Magneschi L, Perata P (2009) Rice germination and seedling growth in the absence of oxygen. Ann Bot 103(2):181–196 ()
41. Magneschi L, Kudahettige RL, Alpi A, Perata P (2009) Comparative analysis of anoxic coleoptile elongation in rice varieties: relationship between coleoptile length and carbohydrate levels, fermentative metabolism and anaerobic gene expression. Plant Biol 11:561–573. doi:10.1111/j.1438-8677.2008.00150.x
42. Magneschi L, Kudahettige RL, Alpi A, Perata P (2009) Comparative analysis of anoxic coleoptile elongation in rice varieties: relationship between coleoptile length and carbohydrate levels, fermentative metabolism and anaerobic gene expression. Plant Biol 11:561–573. doi:10.1111/j.1438-8677. 2008.00150.x
43. Mahender A, Anandan A, Pradhan SK (2015) Early seedling vigour, an imperative trait for direct-seeded rice: an overview on physio-morphological parameters and molecular markers. Planta 241:1027–1050. 10.1007/s00425-015-2273-9 () :. **doi**
44. *Martin DN*, *Proebsting WM*, *Hedden P* (*1999*) *The* SLENDER *gene of pea encodes a gibberellin 2-oxidase. Plant Physiol. 121*:*775–781*
45. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297 – 1297 303 ()
46. Miro B, Ismail AM (2013) **Tolerance of anaerobic conditions caused by flooding during germination and early growth in rice (***Oryza sativa* **L.)**. Front Plant Sci 4:269 ()
47. Miro B, Longkumer T, Entila FD, Kohli A, Ismail AM (2017) Rice seed germination underwater: morpho-physiological responses and the bases of differential expression of alcoholic fermentation enzymes. Front Plant Sci 8:1857. 10.3389/fpls.2017.01857 (). . **doi**
48. *Mohanty B*, *Herath V*, *Wijaya E*, *Yeo HC*, *Reyes BG*, *Lee DY* (*2012*) *Patternsof cis-element enrichment reveal potential regulatory modules involved in the transcriptional regulation of anoxia response of japonica rice*. *Gene 511*:*235–242. doi*:10.1016/j.gene.2012.09.048
49. Nghi KN, Tondelli A, Valè G, Tagliani A, Marè C, Perata P, Pucciariello C (2019) Dissection of Coleoptile Elongation in Japonica Rice Under Submergence Through Integrated Genome-Wide Association Mapping and Transcriptional Analyses. Plant Cell Environ 42:1832–1846. 10.1111/pce.13540) :. **doi**
50. Pearce DME, Jackson MB (1991) **Comparison of growth responses of barnyard grass (***Echinochloa oryzoides***) and rice (***Oryza sativa***) to submergence, ethylene,carbon-dioxide and oxygen shortage**. Ann Bot 68:201–209 ()
51. Perata P, Alpi A (1993) Plant responses to anaerobiosis. Plant Sci 93:1–17 ()
52. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38:904–909 ()
53. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, **de** Bakker PI, 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
54. *Rieu I*, *Eriksson S*, *Powers SJ*, *Gong F*, *Griffiths J*, *Woolley L*, *Benlloch R*, *Nilsson O*, *Thomas SG*, *Hedden P*, *Phillipsa AL* (*2008*) *Genetic analysis reveals that C*19*-GA 2-oxidation is a major gibberellin inactivation pathway in Arabidopsis. Plant Cell 20*:*2420–2436 doi*:10.1105/tpc.108.058818
55. Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26(1):139–140 ()
56. *Ross JJ*, *Reid JB*, *Swain SM*, *Hasan O*, *Poole AT*, *Hedden P*, *Willis CL* (*1995*) *Genetic regulation of gibberellin deactivation in* Pisum. *Plant J. 7*:*513–523*
57. Schadel WE, Walter WM (1981) **Localization of phenols and polyphenol oxidase in "Jewel" sweet potatoes (***Ipomoea batatas "***Jewel"**). Can J Bot 59:1961–1967 ()
58. *Schomburg FM*, *Bizzell CM*, *Lee DJ*, *Zeevaart JAD*, *Amasino RM* (*2003*) *Over expression of a novel class of gibberellin 2-oxidases decreases gibberellin levels and creates dwarf plants*. *Plant Cell 15*:*151–163*
59. Septiningsih EM, Ignacio JC, Sendon PM, Sanchez DL, Ismail AM, Mackill DJ (2013) QTL mapping and confirmation for tolerance of anaerobic conditions during germination derived from the rice landrace Ma-Zhan Red. Theor Appl Genet 126:1357–1366.

doi:10.1007/s00122-013-2057-1

60. Setter T (1994) Relationship between coleoptile elongation and alcoholic fermentation in rice exposed to anoxia. I. importance of treatment conditions and different tissues. Ann Bot 74:265–271. doi:10.1006/anbo.1994.1117
61. Setter TL, Ella ES, Valdez AP (1994) Relationship between coleoptile elonggation difference. Ann Bot 74:273–279 ()
62. Shingaki- Wells R, Millar AH, Whelan J, Narsai R (2014) What Happens To Plant Mitochondria Under Low Oxygen? An Omics Review of The Responses To Low Oxygen And Reoxygenation. Plant Cell Environ 37:2260–2277. doi:10.1111/pce.12312
63. Singh K, Kumar V, Saharawat YS, Gathala M, Ladha JK, Chauhan BS (2013) Weedy rice: an emerging threat for direct-seeded rice production systems in India. J Rice Res 1:106 ()
64. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599. 10.1093/molbev/msm092 () :. **doi**
65. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562–578 ()
66. Wallner S, Dully C, Daniel B, Macheroux P *(*2013*)* Handbook of Flavoproteins, Oxidases, Dehydrogenases and Related Systems (*Hille R., Miller S. M., Palfrey B., editors., eds)Vol.* 1, *De Gruyter, Berlin*
67. *Wang LL, Yu CC, Xu SL, Zhu YG, Huang WC* (*2016*) OsDi19-4 *acts downstream of* OsCDPK14 *to positively regulate ABA response in rice. Plant, Cell & Enviroment 39:2740–2753*
68. Wen ZX, Tan RJ, Yuan JZ, Bales C, Du WY, Zhang SZ, Chilvers MI, Schmidt C, Song QJ, Cregan PB, Wang D (2014) Genome-wide association mapping of quantitative resistance to sudden death syndrome in soybean. BMC Genom 15:809. 10.1186/1471-2164-15-809 (). . **doi**
69. Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ *(*2007*)* An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets. PLoS One 2:e718 *doi:*10.1371/journal.pone.0000718
70. Wu JH, Feng FJ, Lian XM, Teng XY, Wei HB, Yu HH, Xie WB, Yan M, Fan PQ, Li Y, Ma XS, Liu HY, Yu SB, Wang GW, Zhou FS, Luo LJ, Mei HW (2015) Genome-wide association study (GWAS) of mesocotyl elongation based on re-sequencing approach in rice. BMC Plant Biol 15:218. doi:10.1186/s12870-015-0608-0
71. Wu JH, Feng FJ, Lian XM, Teng XY, Wei HB, Yu HH, Xie WB, Yan M, Fan PQ, Li Y, Ma XS, Liu HY, Yu SB, Wang GW, ZhouFS, Luo LJ, Mei HW (2015) Genome-wide association study (GWAS) of mesocotyl elongation based on re-sequencing approach in rice. BMC Plant Biol 15:218 **doi**. 10.1186/s12870-015-0608-0 ()
72. Yamaguchi S, Smith MW, Brown RG, Kamiya Y, Sun T (1998) **Phytochrome regulation and differential expression of gibberellin 3beta-hydroxylase genes in germinating** *Arabidopsis* **seeds**. Plant cell 10:2115–2126 ()
73. Yamauchi M, Aguilar AM, Vaughan DA, Seshu DV (1993) **Rice (***Oryza sativa* **L.) germplasm suitable for direct sowing under flooded soil surface**. Euphytica 67:177–184 ()
74. *Yamauchi Y, Takeda-Kamiya N, Hanada A, Ogawa M, Kuwahara A, Seo M, Kamiya Y, Yamaguchi S* (*2007*) *Contribution of gibberellin deactivation by* AtGA2ox2 *to the suppression of germination of dark-imbibed* Arabidopsis thaliana *seeds. Plant Cell Physiol. 48:555–561*
75. Yang J, Sun K, Li DX, Luo LX, Liu YZ, Huang M, Yang GL, Liu H, Wang H, Chen ZQ, Guo T (2019) Identification of stable QTLs and candidate genes involved in anaerobic germination tolerance in rice via high-density genetic mapping and RNA-SEq. BMC Genom 20(1):355. 10.1186/s12864-019-5741-y () (:. **doi**
76. Yang W, Guo Z, Huang C, Duan L, Chen G, Jiang N, Fang W, Feng H, Xie W, Lian X, Wang GW, Luo QM, Zhang QF, Liu Q, Xiong LZ (2014) Combining high-throughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. Nat Commun 5:5087. 10.1038/ncomms6087 () :. **doi**
77. Yim HK, Lim MN, Lee SE, Lim J, Lee Y, Hwang YS (2012) Hexokinase-Mediated Sugar Signaling Controls Expression Of The Calcineurin B-Like Interacting Protein Kinase 15 Gene and is Perturbed By Oxidative Phosphorylation Inhibition. J Plant Physiol 169:1551–1558. doi:10.1016/j.jplph.2012.06.003
78. Yu HH, Xie WB, Li J, Zhou FS, Zhang QF (2014) A whole-genome SNP array (RICE6K) for genomic breeding in rice. Plant Biotechnol J 12:28–37. doi:10.1111/pbi.12113
79. Zhang MC, Lu Q, Wu W, Niu XJ, Wang CH, Feng Y, Xu Q, Wang S, YuanXP, Yu HY, Wang YP, Wei XH (2017) Association mapping reveals novel genetic loci contributing to flooding tolerance during germination in indica rice. Front Plant Sci 8:678. doi:10.3389/fpls.2017.00678

80. Zhang P, Zhong KZ, Zhong ZZ, Tong HH (2019) **Genome-wide association study of important agronomic traits within a core collection of rice (*Oryza sativa* L.)**. BMC Plant Biol 19:259. 10.1186/s12870-019-1842-7 **()**. . **doi**
81. Zhao KY, Tung CW, Eizenga GC, Wright MH, Liakat Ali M, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. Nat Commun 2:467. doi:10.1038/ncomms1467

Figures

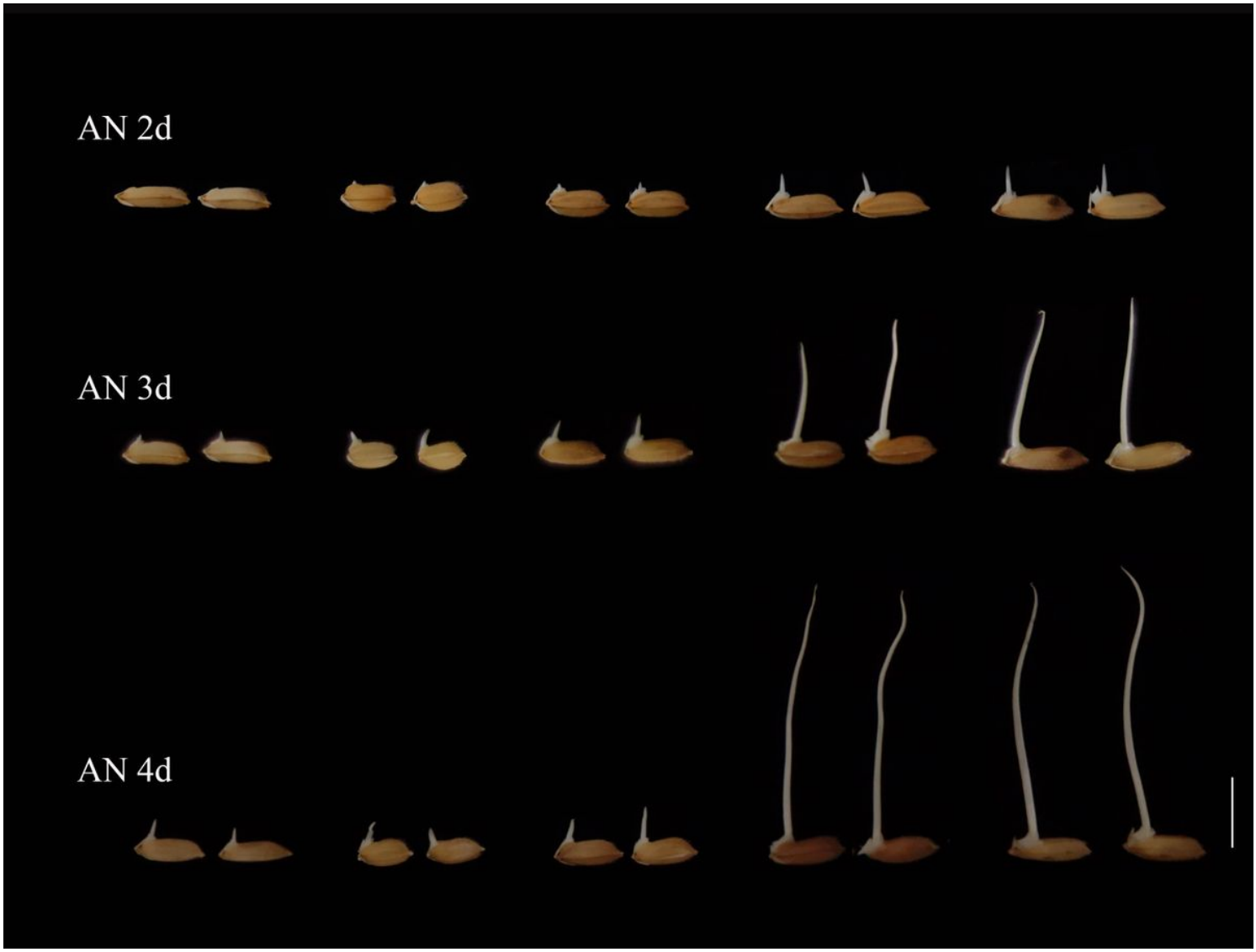

Figure 1

Displaying of the phenotypic of rice coleoptiles during anaerobic germination. Scale = 1 cm

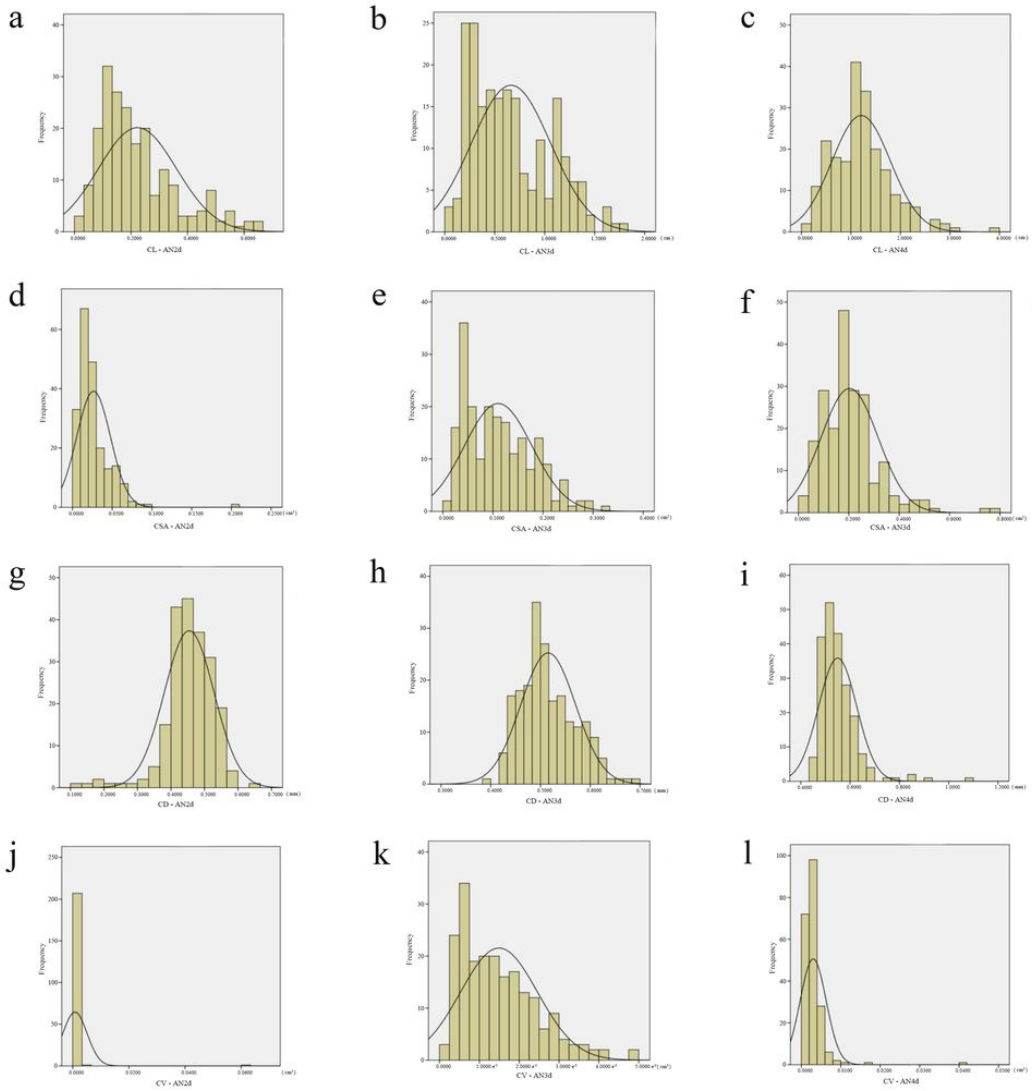

Figure 2

Frequency distribution histogram of each phenotype of coleoptile at different submergence time points. Note: CL, coleoptile length; CSA, coleoptile surface area; CD, coleoptile diameter; CV, coleoptile volume.

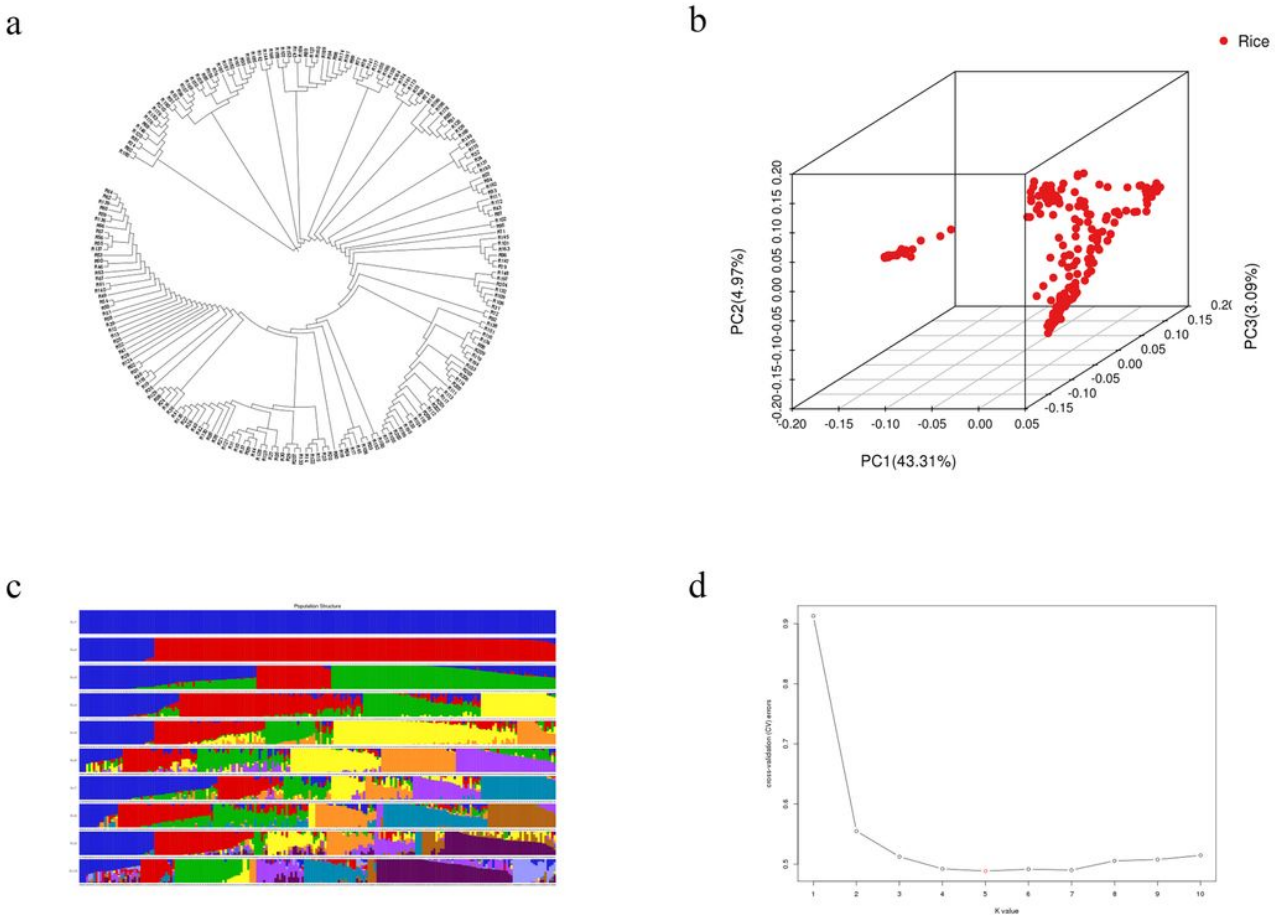

Figure 3

Genetic evolution of natural populations of indica and japonica rice. a, Phylogenetic tree, each branch is a rice germplasm. b, Principal component analysis on 2.12 million SNPs of 209 rice accession. PC 1 and PC 2 refer to the first and second principal components, respectively. The numbers in parentheses refer to the proportion of variance explained by the corresponding axes. Red points represent each variety in 209 rice accession. Each point represented a rice germplasm accession. The closer the distance between the points, the closer the relationship was.c, Cluster analysis results of population genotypes, in which each color represents a group and each row represents a group value. d, Cross validation error rate for each k value. Among them, K is the smallest when k is 5.

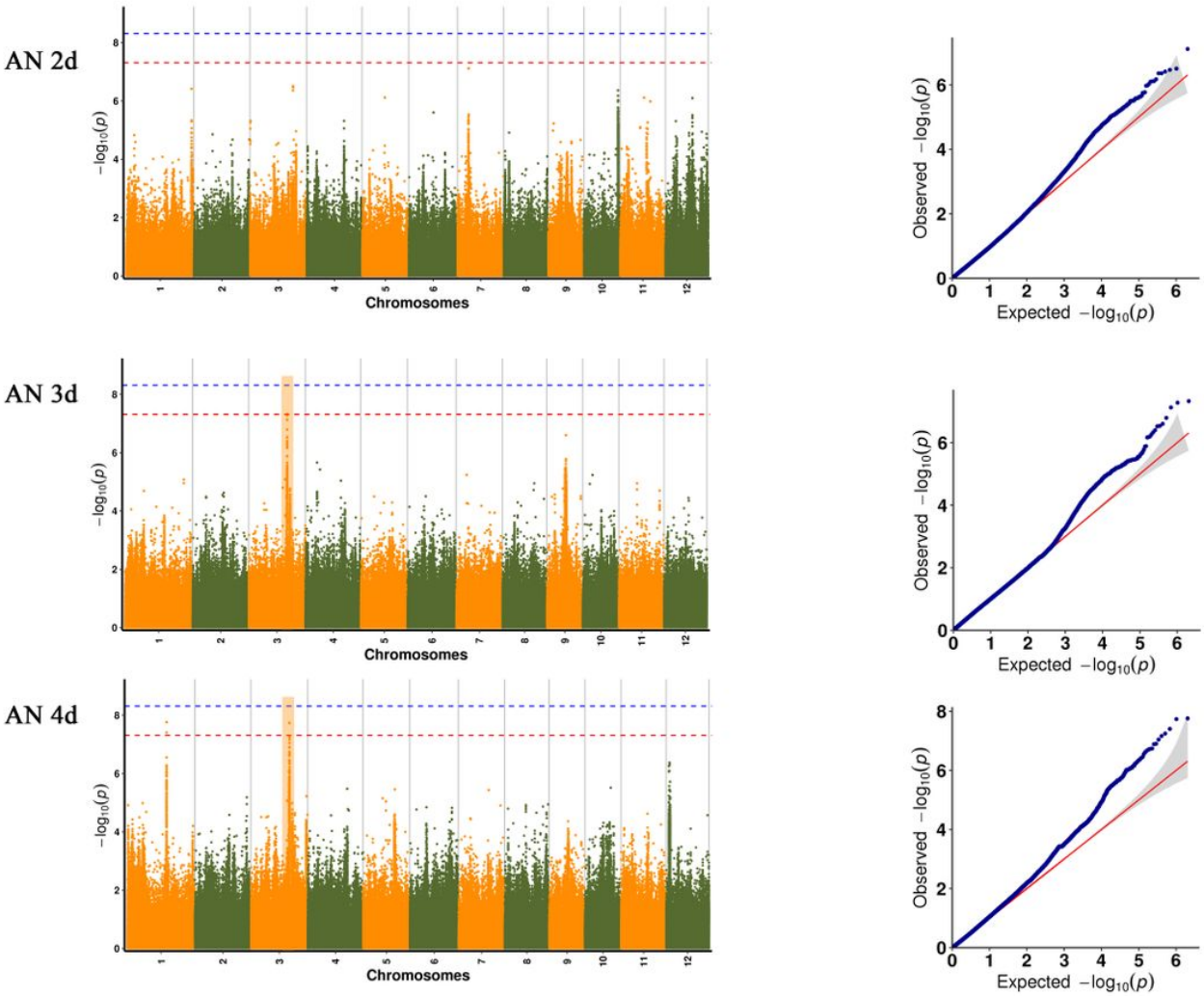

Figure 4

Manhattan map of SNPs associated with coleoptile length at different time points under anaerobic germination. The red horizontal line indicates the significance threshold of multiple comparisons (P < 0.1). The genomic regions detected in our GWAS and previous parental mapping studies were labeled as published QTLs with blue text. Orange blocks represented two co-localized genomic regions under anoxic pressure for different lengths of time in our GWAS study.

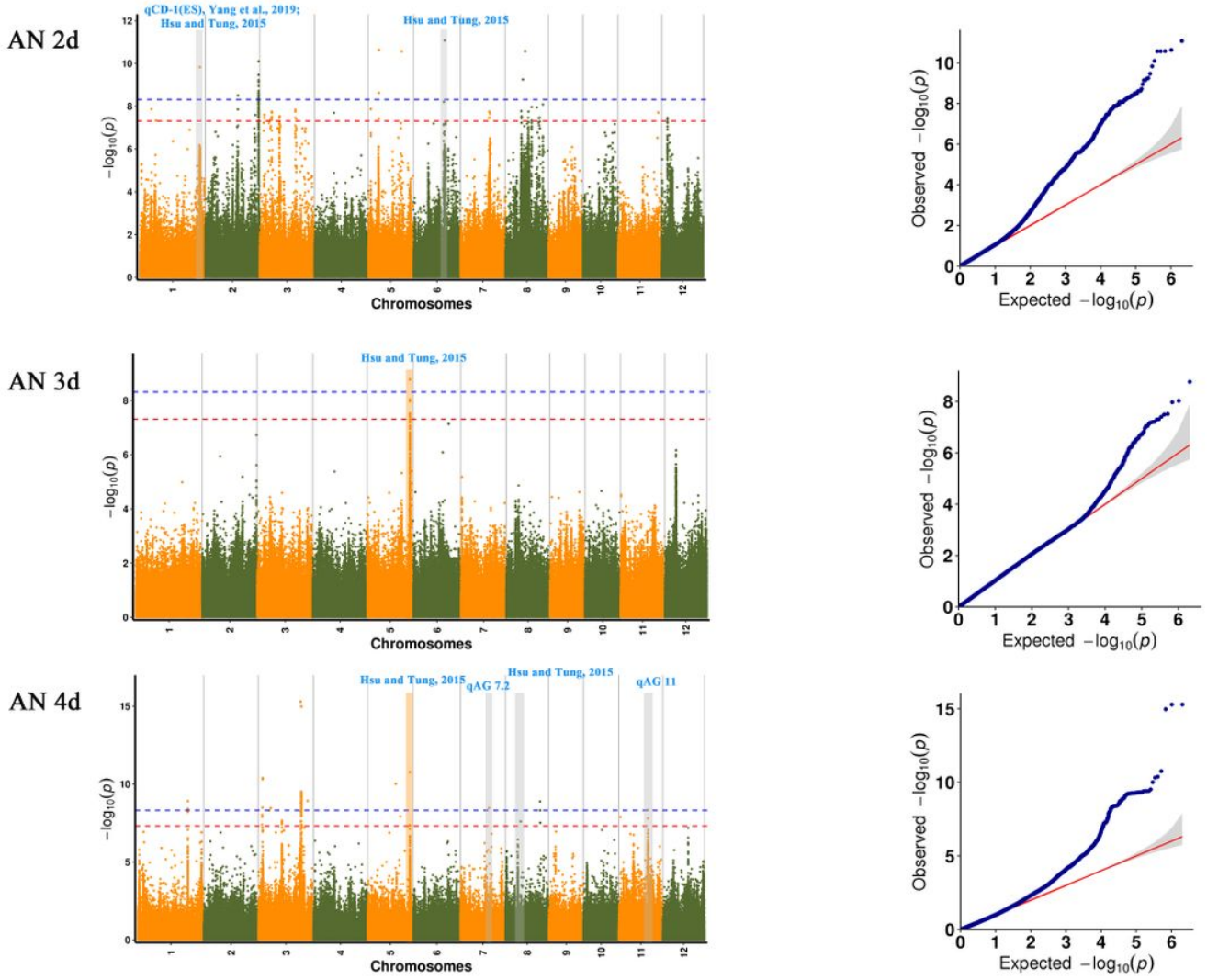

Figure 5

Manhattan map of SNPs associated with coleoptile diameter at different time points under anaerobic germination. The red horizontal line indicates the significance threshold of multiple comparisons (P < 0.1). The genomic regions detected in our GWAS and previous parental mapping studies were labeled as published QTLs with blue text. Orange blocks represented two co-localized genomic regions under anoxic pressure for different lengths of time in our GWAS study.

Gene Id	LOG - Fiod Change			
	AN3dA1dAN3d	AN3dA1dAN4d	A3dAN1dA3d	A3dAN1dA4d
Os03g0639700	0.2892	0.0201	-0.3472	-0.2237
Os03g0639800	0.1773	0.1137	-0.0904	-0.1987
Os03g0640000	-0.8566	-0.5327	-0.6968	0.4756
Os03g0640200	0.1585	0.0888	0.3791	0.1665
Os03g0640300	-0.1174	-0.3033	-0.0121	-0.0815
Os03g0640400	0.1857	0.3403	0.6874	0.2550
Os03g0640800	0.4196	0.4090	-1.0909	-0.9305
Os03g0641200	2.2616	2.5297	-3.7027	-3.3074
Os03g0641250	2.4207	2.0364	-2.5248	-2.1584
Os03g0641300	-0.7146	0.4274	-1.7655	1.0000
Os03g0641500	1.8074	0.3793	0.0000	0.0000
Os03g0641700	0.2856	0.3565	-0.3832	-0.4020
Os03g0642200	0.1664	0.3207	-0.6779	-0.0416
Os03g0642450	1.4263	0.4039	-1.7342	-1.0544
Os03g0642900	-0.8172	-0.8257	0.3051	0.5874
Os03g0559350	1.0583	0.8256	-1.1366	-1.2032
Os03g0559400	-0.0742	-0.3690	0.5580	-0.1043
Os03g0559600	0.4368	0.3594	-0.6601	-0.7381
Os03g0559800	-2.0338	-1.0536	1.3823	0.4854
Os03g0559900	0.0565	0.0846	-0.5411	-0.4531
Os03g0560000	1.6338	1.4006	0.3116	0.6939
Os03g0560100	0.4627	1.5502	0.4081	-0.2432
Os03g0560200	0.4803	0.9068	-0.4730	-0.4116
Os03g0560300	-0.8197	-0.5622	0.1601	0.2040
Os03g0560400	1.5696	1.2489	-2.1841	-2.5081
Os03g0560500	-0.0936	-1.3453	-3.7700	-4.4652
Os03g0560600	0.0833	-0.1799	-1.8522	-2.2596
Os03g0560750	-1.1772	-1.5904	0.0000	0.0000
Os03g0560900	-2.0086	-1.6093	5.1270	2.9412
Os03g0561400	0.0000	0.0000	1.0291	0.0000
Os03g0561500	-0.7192	-0.7463	0.8469	0.6161
Os03g0561600	-0.9474	-0.7196	-0.4208	-0.0354
Os03g0562000	-1.6298	-1.4637	2.8133	1.5210
Os03g0562200	-1.2013	-1.3680	2.0192	1.4208
Os03g0562300	-0.3264	-0.1624	2.1962	1.2252
Os03g0562350	-0.3246	0.3263	1.7213	0.4453
Os03g0562400	0.1088	0.0556	0.0327	-0.0406
Os03g0562800	-1.2185	-1.1693	-1.0702	-0.4698
Os03g0563000	0.8331	1.0841	-3.3995	-2.4754
Os03g0563050	-0.2029	-0.1195	0.5931	0.3344
Os03g0563100	0.3290	0.4815	1.1219	0.7978
Os01g0908800	-0.8106	-0.6277	0.2956	0.2182
Os01g0909000	-0.7232	-0.5146	-0.8991	-1.3451
Os01g0909100	0.9001	0.5581	-0.2267	-0.5171
Os01g0909200	1.0484	1.2094	-1.5375	-1.8357
Os01g0909300	3.0000	3.1926	0.0000	0.0000
Os01g0909400	1.2854	1.1010	-2.1036	-1.3444
Os01g0909500	0.2061	0.4460	0.3143	0.1428
Os01g0910200	0.8197	1.5654	-1.2034	-1.3353
Os01g0910300	-0.2867	-0.3811	1.4001	0.9360
Os01g0909400	0.0000	1.3410	0.0000	0.0000
Os01g0910500	0.4063	-0.1421	-0.5421	-0.2430
Os01g0910800	-1.7500	-1.7500	-1.4463	-2.8093
Os01g0910900	-0.1709	0.8759	0.3447	-0.6101
Os01g0911000	0.2999	0.4564	-0.1306	-0.2988
Os01g0911100	-0.7191	-0.6846	1.0572	0.5546
Os01g0911200	0.5733	0.5101	-1.0000	-0.9636
Os01g0911300	-0.6796	-0.5433	0.3755	0.2199
Os01g0911700	-0.8977	-1.0399	1.4621	1.1708
Os01g0911800	0.1197	0.2424	0.0938	-0.1536
Os01g0911900	0.8559	0.4947	-0.1777	-0.3530
Os06g0547400	0.7524	0.6162	0.6937	0.6964
Os06g0547900	1.3072	1.3868	-1.6828	-1.4581
Os06g0548000	0.4804	0.3770	1.2336	0.8700
Os06g0548100	1.1213	1.2118	0.5568	0.8152
Os06g0548200	-0.7344	-1.0782	3.6078	1.6578
Os06g0549300	2.0268	2.3363	-1.7971	-1.8619
Os06g0549600	0.1855	-0.1427	-0.6911	-1.1854
Os06g0549833	-0.2224	0.0000	0.0000	0.0000
Os06g0549700	0.1460	0.6477	0.4983	0.4857
Os06g0549900	1.3833	1.7217	-0.0238	-0.4034
Os06g0550000	1.7111	1.6300	-1.6283	-0.7527
Os06g0550800	-0.3242	-0.3649	-0.3881	-0.3771
Os06g0551000	-0.0993	0.0000	0.0000	0.0000
Os06g0551400	0.0489	-0.0886	0.1141	0.1149
Os07g0496900	0.1673	-0.2183	-0.3382	-0.0898
Os07g0497000	0.1766	0.0894	-0.0507	-0.0503
Os07g0497100	0.2405	0.4347	0.5809	-0.2552
Os07g0497400	0.7096	0.8943	-0.6647	-0.9012
Os07g0497500	-1.3904	-1.6896	0.0050	0.0301
Os07g0498300	1.4011	1.5269	-1.2043	-0.8328
Os07g0498400	3.7058	3.5986	-3.0682	-2.7013
Os07g0498800	0.3802	0.4229	-0.4128	-0.6102
Os07g0498900	0.4099	0.4559	-1.0831	-1.2216
Os07g0499200	1.1755	0.9053	-2.3532	-2.3703
Os07g0499500	3.6830	0.0000	0.0000	0.0000
Os07g0499800	-1.0114	-0.6849	0.0316	0.1407
Os08g0382500	0.0000	0.0000	-2.3520	-4.4573
Os12g0525600	0.5643	0.6911	-0.5307	-0.4974
Os12g0525700	1.2506	1.4736	-1.0489	-1.3231
Os12g0525800	0.7771	1.4108	-0.5061	-0.3941
Os12g0525900	-2.3821	-2.7747	-0.7157	-2.2118
Os12g0526000	-0.2630	-0.0704	0.0000	0.0000
Os12g0526200	1.6012	1.6969	-1.4945	-0.9011
Os12g0526800	0.3758	0.5525	-0.3061	-0.2678
Os12g0527000	-0.9118	-0.7924	1.1293	0.3392
Os12g0527100	-0.8095	-0.6866	1.2928	1.8777
Os12g0527250	0.9388	1.6491	-1.1220	-1.5850
Os12g0527300	0.3879	0.8540	-1.5209	-1.3279
Os12g0528200	0.6612	0.4700	-1.6143	-1.5621
Os12g0528300	1.5355	1.4784	-0.6692	-0.5394
Os12g0528400	2.0815	3.4203	-0.2575	-0.8699
Os12g0528500	3.2446	3.3320	-2.3520	-1.2940
Os12g0528700	3.9153	0.0000	-3.8349	-3.2869
Os12g0529350	0.0000	0.0000	0.0000	2.6049
Os12g0529900	0.0000	0.0000	-1.0275	-1.0544

Figure 6

Heat map of the fold changes of the 106 candidate differentially expressed genes (DEGs).

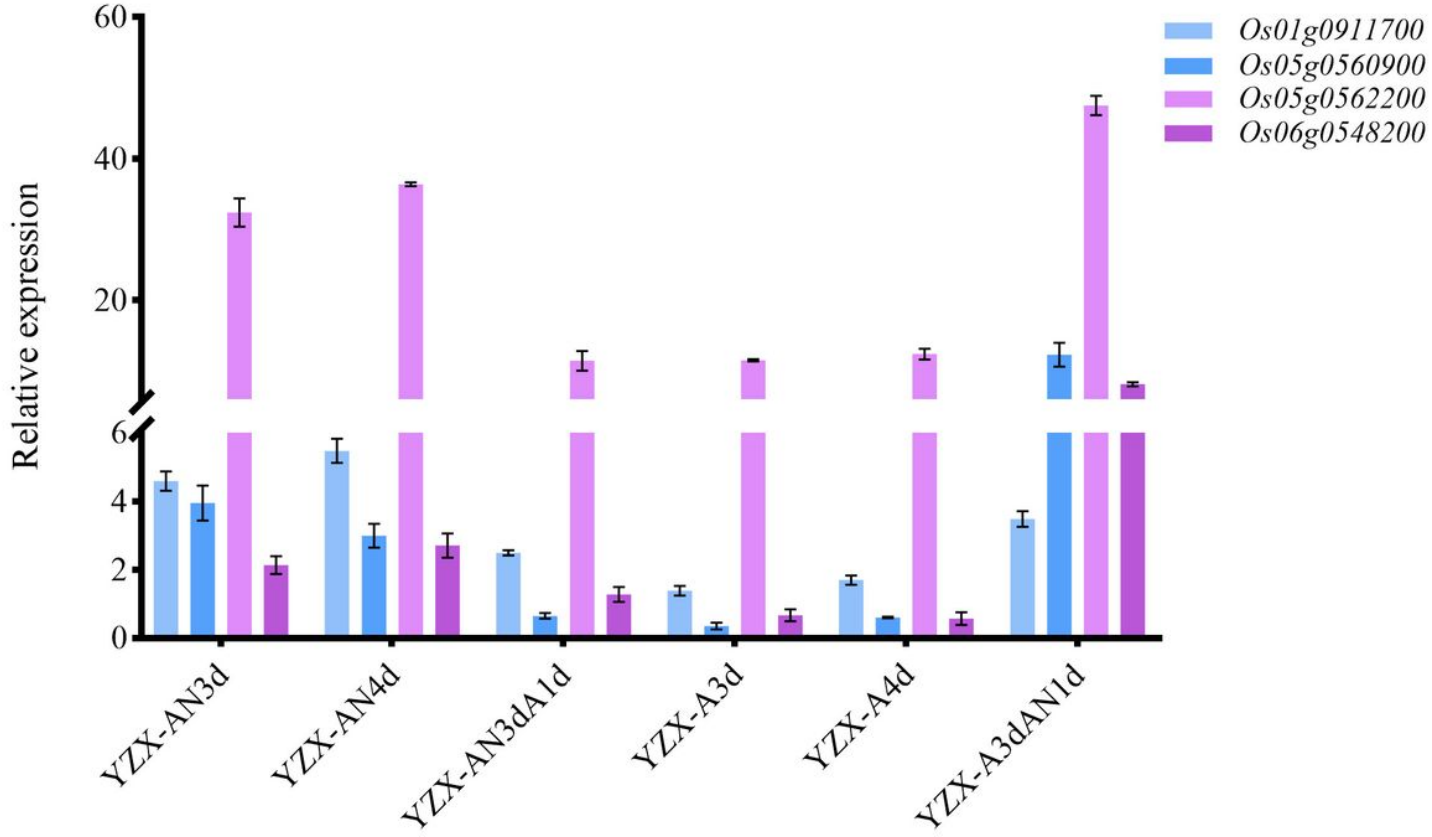

Figure 7

Expression profiles of the four most promising candidate genes via qRT-PCR.

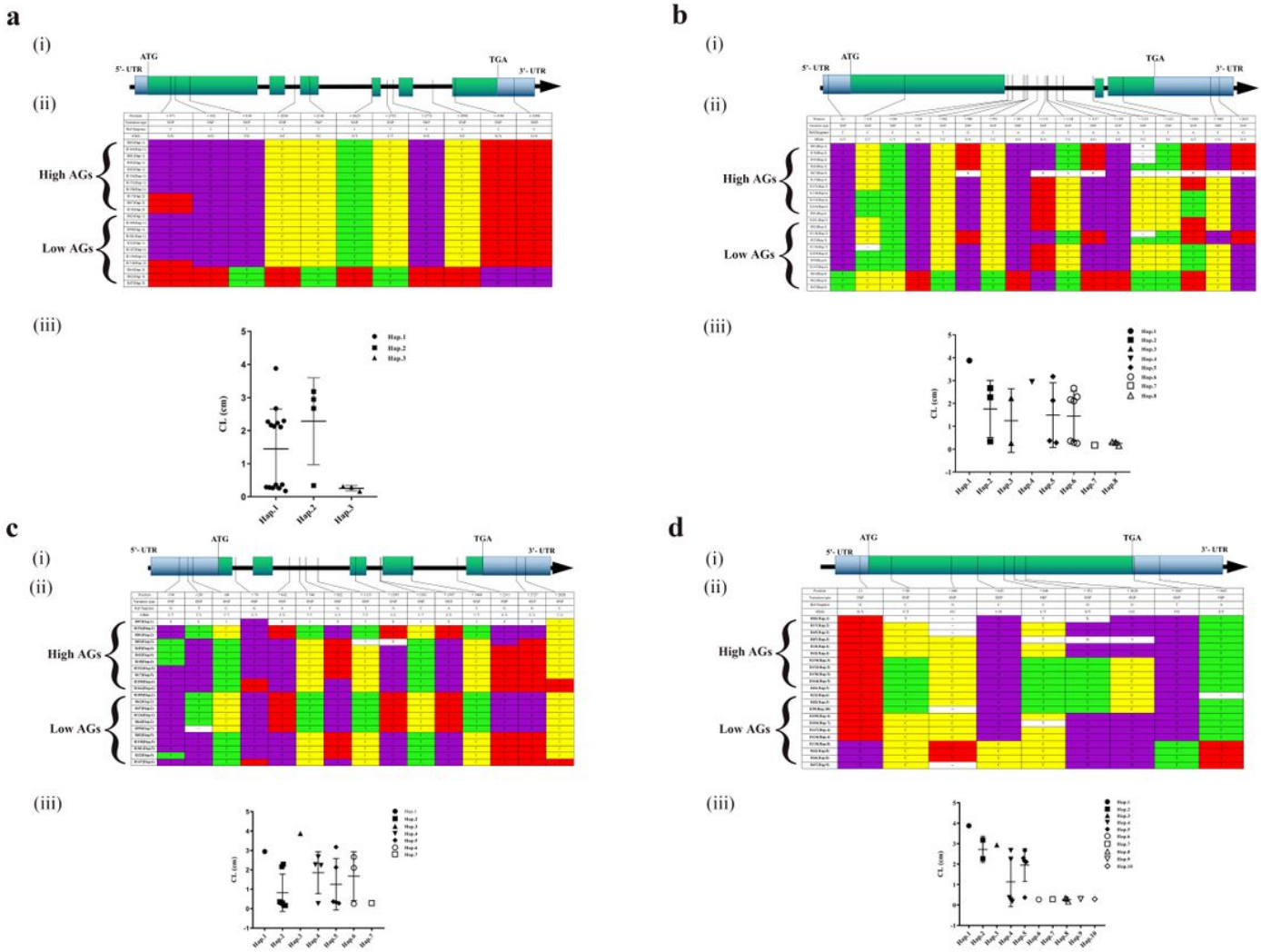

Figure 8

Sequence analysis of four candidate genes in rice with high and low AGs. a. The positions of four candidate genes Os01g0911700、Os05g0560900、Os05g0562200 and Os06g0548200 on chromosomes. b, Sequence analysis of candidate gene Os01g0911700. c, Sequence analysis of candidate gene Os05g0560900. d, Sequence analysis of candidate gene Os05g0562200. e, Sequence analysis of candidate gene Os06g0548200. Notes: (i) the gene structure of four candidate genes is shown respectively, the blue region represents 5' or 3' UTR, the green region represents exon, the straight line represents intron, and the arrow shows gene direction. (ii) the detailed sequence changes of four candidate genes in rice materials with very high and very low AGs are shown, where the position number represents the position related to the "ATG" initiation codon; "-" represents the upstream of "ATG"; and "+" represents the downstream of "ATG". "–" means that the gene is missing 1 bp. (iii) The phenotypic values of each haplotype.

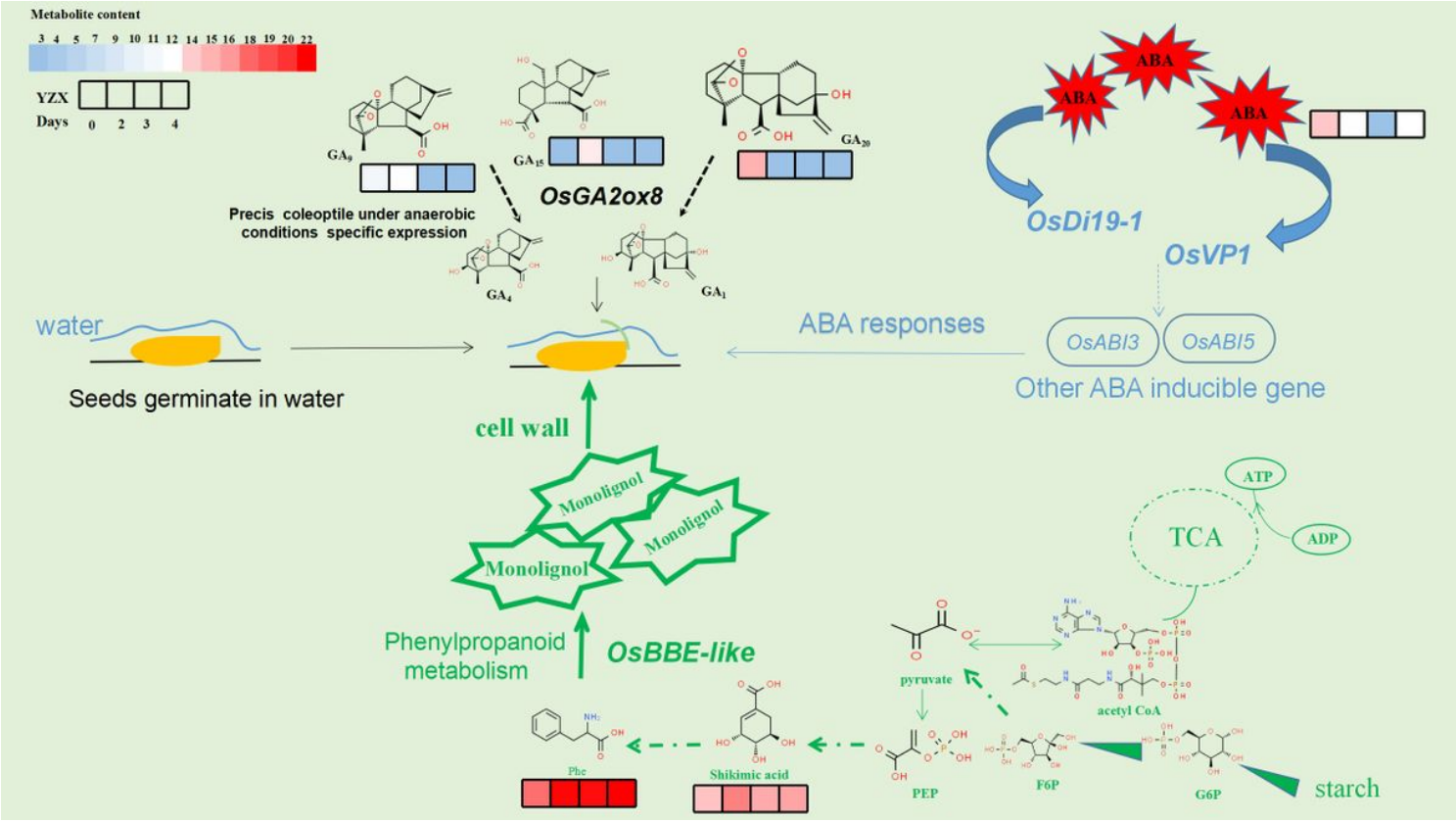

Figure 9

Network map of four candidate genes regulating anaerobic germination of rice seeds. Abbreviations are as follows: G6P, Glucose-6-phosphate; F6P, Fructose-6-phosphate; PEP, phosphoenolpyruvic acid; Phe, phenylalanine; TCA, tricarboxylic acid cycle.

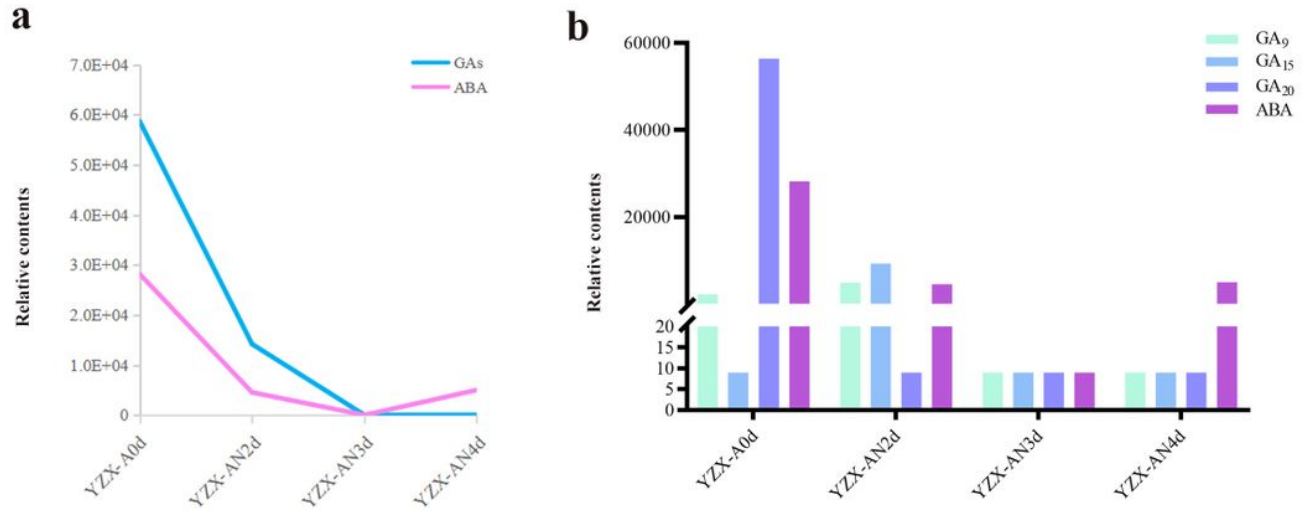

Figure 10

a, The content of GAs and ABA in the condition anaerobic germination.b, The contents of three kinds of GA: GA9/GA15/GA20 and ABA under the condition of anaerobic germination.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- supplement17.tif
- supplement18.tif
- supplement19.xls
- supplement20.xls
- supplement21.xls
- supplement22.xls
- supplement23.xls