Theoretical and Applied Genetics
https://doi.org/10.1007/s00122-022-04065-4

ORIGINAL ARTICLE

GWAS and transcriptome analysis reveal *MADS26* involved in seed germination ability in maize

Langlang Ma[1] · Chen Wang[1] · Yu Hu[2] · Wei Dai[1] · Zhenjuan Liang[1] · Chaoying Zou[1] · Guangtang Pan[1] · Thomas Lübberstedt[3] · Yaou Shen[1]

Received: 22 November 2021 / Accepted: 15 February 2022

Abstract

Key message ***MADS26* affecting maize seed germination was identified by GWAS and transcriptomics. Gene-based association analyses revealed three variations within *MADS26* regulating seed germination traits. Overexpressed *MADS26* in *Arabidopsis* improved seed germination.**

Abstract Seed germination ability is extremely important for maize production. Exploring the genetic control of seed germination ability is useful for improving maize yield. In this study, a genome-wide association study (GWAS) was conducted to excavate the significant SNPs involved in seed germination ability based on an association panel consisting of 300 lines. A total of 11 SNPs and 75 candidate genes were significantly associated with the seed germination traits. In addition, we constructed 24 transcriptome libraries from maize seeds at four germination stages using two inbred lines with contrasting germination rates. In total, 15,865 differentially expressed genes were induced during seed germination. Integrating the results of GWAS and transcriptome analysis uncovered four prioritized genes underlying maize seed germination. The variations located in the promoter of *Zm00001d017932*, a *MADS-transcription factor* 26 (*MADS26*), were verified to affect the seed germination, and the haplotype TAT was determined as a favorable haplotype for high-germination capability. *MADS26* was induced to express by ethylene during seed germination in maize and overexpressing *MADS26* increased the seed germination ability in *Arabidopsis*. These findings will contribute to understanding of the genetic and molecular mechanisms on seed germination and the genetic modification of seed germination ability in maize.

Introduction

As an important agronomic trait, seed germination ability influences the vegetative growth and yield formation of crops, which is controlled by genetic and environmental factors (Han et al. 2014). Seeds with high-germination ability significantly improve the crop gain relative to those with low-germination ability (Hu et al. 2016). Maize (*Zea mays* L.) is widely grown and hugely demanded for humans and animals, worldwide (Ma et al. 2018); however, the maize production is frequently compromised by the low seed germination ability (Hu et al. 2016). Currently, the molecular mechanism of seed germination is still obscure in maize. Therefore, decoding the genetic basis and causal genes controlling maize seed germination is an urgent need for improving maize yield.

The process of seed germination is generally divided into three phases: Phase I is the stage of water absorption by mature dry seeds; phase II is characterized by several biosynthesis and metabolism processes, such as protein synthesis, DNA repair, and mitochondrial activity, with the seed water content maintaining relatively high constant; during phase III, the tender radicles emerging at the end of phase II keep water uptake and process cell division and elongation

Communicated by Benjamin Stich.

Langlang Ma and Chen Wang have contributed equally to this work.

✉ Yaou Shen
shenyaou@sicau.edu.cn

[1] State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China, Maize Research Institute, Sichuan Agricultural University, Chengdu 611130, China

[2] Zigong Research Institute of Agricultural Sciences, Zigong 643002, China

[3] Department of Agronomy, Iowa State University, Ames 50010, USA

Published online: 05 March 2022

until the seedling formations (Bewley et al. 2012; Guo et al. 2019).

In the past decades, numerous genetic loci controlling seed germination ability were identified in *Arabidopsis*, rice, wheat, and other species using linkage analysis (Landjeva et al. 2010; Liu et al. 2014; Yuan et al. 2016). A total of three quantitative trait loci (QTL) involved in seed germination speed were identified on chromosomes 1, 3, and 4 in *Arabidopsis* (Yuan et al. 2016). In wheat, ten seed germination-associated QTL were detected using a D genome introgression line population (Landjeva et al. 2010). In rice, seven QTL controlling seed germination rate were mapped by a recombinant inbred line (RIL) population, and three of them were defined as major QTL with phenotypic variation > 10% (Liu et al. 2014). In maize, previous studies mainly focused on the dissection of genetic control for seed germination ability under adverse stresses. Only few QTL were shown to confer seed germination ability under optimum conditions. For example, three QTL *qOTGR5-1*, *qOTGR6-1*, and *qOTGR7-1* were found to be associated with germination rate by using an IBM Syn4 RIL population (Hu et al. 2016).

With the decrease in DNA sequencing cost, the identification of high-density molecular markers on a large scale becomes feasible for researchers. According to the high-density markers, genome-wide association studies (GWAS) have developed into an efficient tool in excavating the causal candidate genes associated with target traits. Using GWAS, Guo et al. (2019) revealed six significant SNP (single nucleotide polymorphism)-trait associations for seed germination rate based on a rice natural panel and 161,657 high-quality SNPs. To our knowledge, GWAS has not been applied in the genetic control dissection of seed germination ability under optimum conditions in maize. As such, the genetic and molecular mechanisms underlying seed germination remain largely unknown in maize. Additionally, GWAS cannot refine the candidate genes, which requires the combination with linkage mapping or transcriptome analysis for pinpointing the trait-associated genes (Yao et al. 2020). By integrating the GWAS results and differentially expressed genes (DEGs), Guo et al. (2020) identified seven priority genes for seminal root length of maize seedlings under drought stress. Similarly, a total of 21 candidate gene involved in *Fusarium* ear rot resistance were co-localized in maize using the above strategy, five of which were situated in a resistance hotpot region (Yao et al. 2020).

Several genes controlling seed germination have been reported in plants, such as *Aldehyde Dehydrogenase 7* (*OsALDH7*) (Shin et al. 2009) and *Lipoxygenases* (*LOXs*) (Suzuki et al. 1996) in rice, and *Lipid-hydrolyzing Phospholipase D* (*PLDa1*) (Devaiah et al. 2007), *Protein-L-isoaspartate O-methyltransferase* (*PIMT1*) (Oge et al. 2008), *Flowering Locus C* (*FLC*) (Chiang et al. 2009), and *Agamous-Like67* (*AGL67*) (Bassel et al. 2011) in *Arabidopsis*. Among these genes, *FLC* and *AGL67* encode the MADS-box transcription factors that contain a highly conserved *N*-terminal DNA binding domain with 55–60 amino acids (Masiero et al. 2011). According to their evolutionary lineage, the plant *MADS-box* transcription factors were classified into two main types, type I and type II genes (Masiero et al. 2011). Compared to the type I genes, the type II genes have more complex gene structures with an additional *K* domain (Tian et al. 2015). Meanwhile, the type II genes could be further grouped into the MIKCC- and MIKC*-types, based upon their corresponding structural features (Henschel et al. 2002). So far, most of the functionally known *MADS-box* transcription factors belong to the MIKCC-type genes (Tian et al. 2015). In plants, the *MADS-box* genes were reported to play key roles in regulating the development of inflorescence and floral meristems, flowering time, root growth, and fruit ripening (Masiero et al. 2011).

In the present study, a GWAS was conducted in a maize association panel to identify the genetic loci and initial candidate genes involved in seed germination ability. We then analyzed the transcriptomes of two lines with contrasting germination ability at different germination stages, detecting the DEGs responsive to seed germination. The candidate genes for seed germination traits were uncovered by a combination of GWAS results and DEGs. The genes were each PCR-amplified in 68 inbred lines to perform gene-based association studies, identifying the germination-associated variations in the potential causal gene (*Zm00001d017932*). Finally, *Zm00001d017932* that encodes a *MADS-transcription factor 26* (*MADS26*) was overexpressed in *Arabidopsis* for functional validation. Our goals were to (1) improve the understanding of the genetic control of seed germination ability in maize, (2) identify the causal gene and its favorable haplotype affecting seed germination ability, and (3) accelerate the application of marker-assisted breeding for seed germination ability.

Materials and methods

Plant materials and seed germination trials

An association panel consisting of 300 inbred lines was subjected to evaluation of the seed germination ability. The 300 inbred lines were collected from Southwest China breeding program, which included tropical, non-stiff stalk (NSS), stiff stalk (SS), and other germplasms (Zhang et al. 2016). This maize panel was planted in Chongzhou (E103° 67′, N30° 63′) of Sichuan Province, and the seeds from each line were harvested at the stage of physiological maturity and oven-dried at 37 ℃.

The germination trials were performed in a cultivation box using a randomized complete block design with three

biological repetitions. The parameters of cultivation box were set as: light/darkness = 16/8 h; temperature under light/temperature under darkness = 28/24 ℃; relative humidity = 65%. For each repetition, after removing the shriveled or unfilled seeds, 30 healthy and uniform-sized seeds were treated with 10% H_2O_2 for 30 min and then rinsed three times using distilled water (Zhang et al. 2021). The sterilized kernels were soaked with saturated $CaSO_4$ $2H_2O$ for 6 h and then washed three times by distilled water (Zhang et al. 2021). The seeds were placed in a Petri dish (size: 13 cm × 13 cm) with single-layered filter paper on the bottom for germination (Zhang et al. 2021).

Phenotypic data collection and statistical analysis

Herein, the seed germination ability was decomposed into five traits, namely seed germination rate on 3 d (SGT), seed germination rate on 7 d (SGS), root length on 7 d (RLS), shoot length on 7 d (SLS), and root–shoot ratio on 7 d (RSRS). The seed germination rate was equal to germinated seed number divided by total seed number. Root–shoot ratio was calculated as root length divided by shoot length. The shoot length and root length were measured by a ruler.

The SPSS (Statistical Product and Service Solutions, version 21.0, IBM, Armonk, NY) tool was utilized to calculate the mean, minimum (Min), maximum (Max), and standard deviation (SD) values. Distributions and correlations of trait phenotypes were analyzed by using a "PerformanceAnalytics" package in the RStudio software. For each trait, the broad-sense heritability (H_B^2) was estimated by the formula: $H_B^2 = \sigma_G^2/\sigma_P^2$, where σ_G^2 and σ_P^2 represent genotype variance and phenotype variance, respectively; σ_G^2 and σ_P^2 were calculated as follows: $\sigma_G^2 = (MSG\text{-}MSE)/rep$, $\sigma_P^2 = ((MSG\text{-}MSE)/rep) + MSE$, where *MSG*, *MSE*, and *rep* represent the mean square of genotype, mean square of error, and biological repetition numbers, respectively (Pace et al. 2015). The indicators *MSG* and *MSE* were calculated using a SAS 9.3 (Statistical Analysis System, version 9.3, SAS Institute, Cary, NC, USA) software.

Genome-wide association study

In our previous study, 56,110 SNPs were generated in this association panel by the Illumina MaizeSNP50 BeadChip (Zhang et al. 2016). According to the filtering principles of missing rate > 20%, heterozygosity > 20%, or minor allele frequency (MAF) < 0.05, a total of 43,675 high-quality SNPs were retained for GWAS in this study. Previous results revealed that the 300 germplasms of this association panel can be divided into three subpopulations and the linkage disequilibrium (LD) decay was approximately 220 kb at the $r^2 = 0.2$ (Zhang et al. 2016, 2021; Ma et al. 2018). Thus, as a covariate, the value of principal component analysis (PCA) calculated by GAPIT software (Lipka et al. 2012) was added to the fixed and random model circulating probability unification (FarmCPU) model for GWAS (Liu et al. 2016). The simpleM program in R (Cao et al. 2010; Johnson et al. 2010) was adopted to calculate the effective number of the 43,675 SNPs. In total, 24,390 independently effective SNPs were ultimately obtained and the threshold of significant SNP-trait association was set as: $P = 0.05/N$ (N is the effective SNP number) = 2.05×10^{-6}. The gene models situated within the LD region of each significant SNP were considered as the initial candidate genes responsible for the corresponding trait.

RNA-seq and data analysis

Two inbred lines SCL326 (high seed germination rate) and SCL127 (low seed germination rate) were selected from the association panel for transcriptomic analysis. Briefly, the seeds of SCL326 and SCL127 were germinated under optimum conditions, as described in the germination experiments of the association panel. For each line, six seeds were collected at 0, 12, 24, and 48 h of germination, respectively, with three biological repetitions. Total RNA was extracted from the mixed samples of the six seeds at each stage. In total, 24 RNA samples from the two lines were sequenced by Illumina HiSeq TM 2000 platform (San Diego, CA). Clean reads were obtained from the filtration and quality control of original sequencing data using fastp software (Chen et al. 2018b). The clean reads were then aligned to maize genome (ZmB73 RefGen_V4) via Hisat2. The FPKM (Fragments Per Kilobase Million) was utilized to normalize and estimate gene expression values. DEGs between two pairwise samples were revealed by DESeq (Anders and Huber 2010), with the criteria as follows: $|\log_2$ fold-change (FC)$| \geq 2$ and adjusted $P < 0.05$. Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed using a OmicShare platform (https://www.omicshare.com).

Candidate gene association study

Four genes (*Zm00001d017932*, *Zm00001d003657*, *Zm00001d029793*, and *Zm00001d017906*) obtained from the combination of GWAS and transcriptome analysis were separately subjected to a gene-based association study. For each gene, the gene body and promoter (upstream 2000 bp) regions were PCR-amplified among 68 inbred lines randomly selected from the association panel. DNAMAN program (Version5.2.2, Lynnon Bio-soft, Canada) was used to uncover the nucleotide polymorphisms in each gene among the 68 inbred lines based on B73 RefGen_v4 genome (Jiao et al. 2017; Li et al. 2020). The phenotypic data and genetic variations with MAF > 0.05 were combined

to conduct association analysis using the general linear model (GLM) in TASSEL 5.0 software. The variation loci with a P value < 0.05 were considered significantly associated with the target traits (Ma et al. 2021). For each gene, the significant variations were used for haplotype division, and the phenotypic difference between haplotypes was analyzed via a t test. HaploView software package was used to calculate the LD decay between the pairwise nucleotide polymorphisms in each gene (http://www.broad.mit.edu/mpg/haploview/).

Zm00001d017932 expression analysis

Candidate gene association studies indicated that only the significant markers within *Zm00001d017932* affected seed germination traits. As such, *Zm00001d017932* was regarded as a high-priority candidate gene underlying seed germination ability in maize.

To identify the expression patterns of *Zm00001d017932*, we selected three haplotype II-lines (SCL016, SCL066, and SCL231) with high SGS and three haplotype III-lines (SCL023, SCL091, and SCL142) with low SGS for quantitative real-time PCR (qRT-PCR). For each material, RNA samples were isolated from the seeds at 0, 12, 24, and 48 h of germination using TRIZOL reagent (Invitrogen) and cDNAs were synthesized using PrimeScript RT Reagent Kit With gDNA Eraser (TaKaRa). Primer 5.0 software was used to design the primer pair of *Zm00001d017932* (Table S1). Gene relative expression was calculated by using $2^{-\Delta\Delta Ct}$ method, with *ACTIN 1* (*Zm00001d010159*) as the internal control (Table S1). The qRT-PCR amplifications for each sample were conducted with three biological repetitions.

The promoter sequences of *Zm00001d017932* were cloned from a haplotype II-line (SCL231) and a haplotype III-line (SCL142) and inserted the upstream of *GUS* on the vector pCAMBIA3301, respectively. The two constructs both contained a *GFP* in the T-DNA region under the control of *35S* promoter, which was used to normalize the *GUS* expression levels. Subsequently, the two recombinant vectors were each transformed into *Nicotiana tabacum* leaves via *Agrobacterium* injection. The relative *GUS* transcript level (the ratio between *GUS* and *GFP* transcript abundances) was determined in the two haplotypes using qRT-PCR with the primers listed in Table S2 (Yoo et al. 2007). *NtActin* acted as the internal reference gene in the qRT-PCR.

Zm00001d01793 encodes a MADS transcription factor, *MADS26*. Previous studies demonstrated that the MADS transcription factors participated in plant growth and development via ethylene biosynthesis, ethylene perception, and downstream ethylene response (Martel et al. 2011; Fujisawa et al. 2013). To confirm the response of *Zm00001d01793* to ethylene, we treated the seeds of three haplotype III-lines (SCL023, SCL091, and SCL142) and three haplotype II-lines (SCL016, SCL066, and SCL231) with 400 mg/L ethylene and distilled water (control), respectively. On the 4, 5, 6, and 7 d of germination, the shoot and root parts of each line were individually collected for qRT-PCR. Meanwhile, we also investigated the SGS of the three haplotype II-lines under the ethylene treatment.

Subcellular localization of MADS26

The coding sequence (CDS) of *MADS26* was amplified from B73 and inserted into the pCAMBIA2300-35S-eGFP vector to generate the MADS26-eGFP fusion construct, *p35S: MADS26-eGFP*. The recombinant plasmid was introduced into tobacco leaves via *Agrobacterium* injection. The laser confocal microscope was used to observe the eGFP fluorescence at 48 h of *Agrobacterium* injection. In addition, we also transformed the recombinant plasmid into maize mesophyll protoplasts via a PEG4000-mediated method. After 48 h incubation, the eGFP fluorescence was examined under a laser confocal microscope (Yoo et al. 2007; Ren et al. 2017).

Investigation of the seed germination ability in *MADS26*-overexpressed *Arabidopsis*

The *MADS26* CDS was amplified from the high-SGS line (SCL231) and inserted into the expression vector pRI101-AN under the control of 35S promoter for constitutive expression. The recombinant vector *p35S*: $MADS26^{SCL231}$ was transformed into *Agrobacterium tumefaciens* strain GV3101. The transformed *Agrobacterium* was then used to infect the *Arabidopsis thaliana* ecotype *Columbia* (Col) based on the floral dip method (Clough and Bent 1998). Positive homozygotes of the T_3 transgenic events were obtained by kanamycin-resistant selection and *MADS26*-based PCR.

Three *MADS26*-overexpressed transgenic events and the wide-type Col were subjected to investigation of the seed germination traits. The seeds were sterilized and planted in 1/2 MS culture mediums. The germination rate was recorded at 24 h and 48 h of seed germination. The root lengths were measured on the 5th day.

Results

Seed germination ability in the association panel

In this study, five traits RLS, RSRS, SLS, SGS, and SGT were utilized to assess the seed germination ability of each maize line. All the traits presented abundant variations, with the SDs and CVs ranging from 0.29 to 3.66 and 0.36 to 0.52, respectively (Table S3). The H_B^2 estimates for the five traits were in the range of 82.83–93.38%, suggesting

that the seed germination ability was mainly controlled by genetic factors (Table S3). Significantly ($P < 0.001$) positive correlations were observed between each pair of the four traits (RLS, SLS, SGS, and SGT), with the correlation coefficients ranging from 0.27 (SGS and SLS) to 0.91 (SGS and SGT) (Table S4). Furthermore, RSRS and SLS displayed a significantly negative correlation ($r = -0.50$, $P < 0.001$). In contrast, RSRS and RLS showed a significantly positive correlation ($r = 0.56$, $P < 0.001$) (Table S4). These findings indicated that the five traits probably exert a synergistic effect on maize seed germination. In addition, the phenotype frequency distributions of the three traits RLS, RSRS, and SLS followed normal distributions (Fig. S1), suggesting that they were genetically controlled by multiple genes.

Significant loci and candidate genes affecting seed germination ability revealed by GWAS

In GWAS, the GLM and mixed linear model (MLM) separately result in many false positive and negative associations (Hu et al. 2017; Ma et al. 2020). The FarmCPU model addresses the confounding problem of testing SNPs by iterative algorithms (Hu et al. 2017; Ma et al. 2020). As such, we used the FarmCPU model to perform GWAS in the present study. A total of two, two, three, two, and two significant SNPs associated with SGS, RLS, SGT, SLS, and RSRS were identified at a P value threshold of 2.05×10^{-6}, respectively (Fig. 1). The most significant SNP, SYN1938 ($P = 1.07 \times 10^{-8}$), was found to be involved in the trait SGT (Fig. 1). According to the LD decay (220 kb) of this association panel, we identified 17, 11, 30, 14, and 3 unique candidate genes influencing SGS, RLS, SGT, SLS, and RSRS, respectively (Table S5). Referring to their functional annotations, the homologues of seven candidate genes *Zm00001d003657*, *Zm00001d007234*, *Zm00001d017913*, *Zm00001d017932*, *Zm00001d018984*, *Zm00001d019035*, and *Zm00001d042376* were known to mediate the processes of seed germination and development (Table S5).

Differentially expressed genes responding to seed germination

Via transcriptome sequencing for the two lines with contrasting germination abilities, a total of 1.42 billion clean reads were obtained from 24 libraries of the lines. Among them, 1.28 billion clean reads were uniquely mapped to the B73 RefGen_v4 reference genome, with the mapping rate ranged from 87.65 to 93.07% (the average rate: 90.27%) (Fig.S2). The correlation analysis revealed high-quality biological replicates for each sample (Fig. S3a). In the 24 samples, approximately 64.3, 33.3 and 2.4% of the genes were expressed at low ($0 \le FPKM < 1$), moderate ($1 \le FPKM < 60$), and high ($FPKM > 60$) levels, respectively (Fig. S3b).

To pick out the genes involved in seed germination ability, we identified the DEGs between different samples (12 h vs. 0 h, 24 h vs. 0 h, and 48 h vs. 0 h) for each line. Relative to the gene expression levels at 0 h, 1219 (19 upregulated and 1200 downregulated), 3997 (999 upregulated and 2998 downregulated), and 5386 (2431 upregulated and 2955 downregulated) DEGs were detected in the low-germination line SCL127 at 12, 24, and 48 h of seed germination, respectively (Fig. 2a, b). However, 4353 (713 upregulated and 3,640 downregulated), 7106 (3013 upregulated and 4093 downregulated), and 8139 (2865 upregulated and 5274 downregulated) genes were differentially expressed at 12, 24, and 48 h of seed germination in the high-germination line SCL326, respectively (Fig. 2c, d). In addition, we separately obtained 4545, 4379, 5455, and 5127 DEGs between SCL127 and SCL326 at 0, 12, 24, and 48 h during seed germination (Fig. 2e, f). Of these, 2946, 2478, 3898, and 2640 DEGs, respectively, were upregulated in SCL326 when compared to those in SCL127 (Fig. 2e). Combining the above findings, we excavated 15,865 DEGs for seed germination response in total.

To know the functional information of the 15,865 DEGs, we performed the GO enrichment and KEGG analyses. The 15,865 DEGs were significantly enriched in the seed development or seed germination-related GO terms, such as nucleic acid binding activity (GO: 0001071), transferase activity (GO: 0016758), and hydrolase activity (GO: 0,004,553) (Hartweck et al. 2002; Kim et al. 2007; Katsuya-Gaviria et al. 2020; Latha et al. 2021) (Fig. S4). KEGG analysis showed that the top four pathways were metabolism, biosynthesis of secondary metabolite, starch and sucrose metabolism, and plant hormone signal transduction (Fig. S5). Previous studies revealed that these pathways were correlated with seed development or seed germination in plants (Rochat and Boutin 1992; Weber et al. 1997; Kucera et al. 2005).

Priority candidate genes identified by integrating the GWAS and transcriptome data

Among the 75 genes detected by GWAS, 34 were among the 15,865 DEGs identified from transcriptome data (Fig. 3a). Herein, we further focused these candidates in the 34 DEGs, which were specifically differentially expressed in all the pairwise comparisons (12 h vs. 0 h, 24 h vs. 0 h, and 48 h vs. 0 h) of the high-germination line, SCL326. Finally, the four candidate genes, *Zm00001d017932*, *Zm00001d003657*, *Zm00001d029793*, and *Zm00001d017906,* were considered specifically responsive to the whole seed germination process in SCL326 (Fig. 3a). These genes separately encode an agamous-like MADS-box protein, a scarecrow protein, a glutathione *S*-transferase F9, and an *O*-fucosyltransferase family protein (Table S5). Among them, the three genes

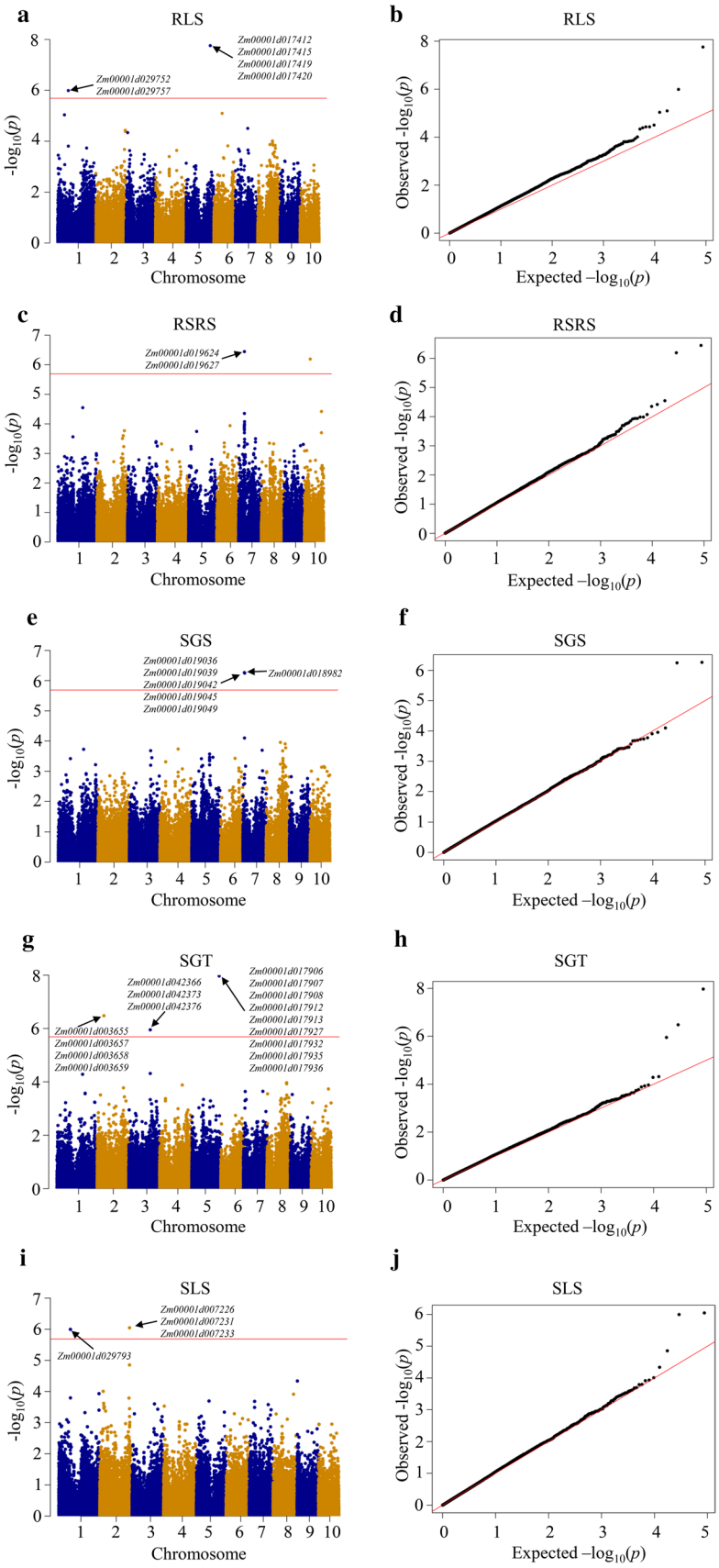

Fig. 1 Significant SNPs detected in this study. **a**, **c**, **e**, **g** and **i** Manhattan plots of the significant SNPs associated with RLS, RSRS, SGS, SGT, and SLS, respectively. **b**, **d**, **f**, **h** and **j** Q–Q (Quantile–Quantile) plots of the significant SNPs associated with RLS, RSRS, SGS, SGT, and SLS, respectively. SNPs, single nucleotide polymorphisms. SGT, seed germination rate on 3 d; SGS, seed germination rate on 7 d; RLS, root length on 7 d; SLS, shoot length on 7 d; RSRS, root–shoot ratio on 7 d

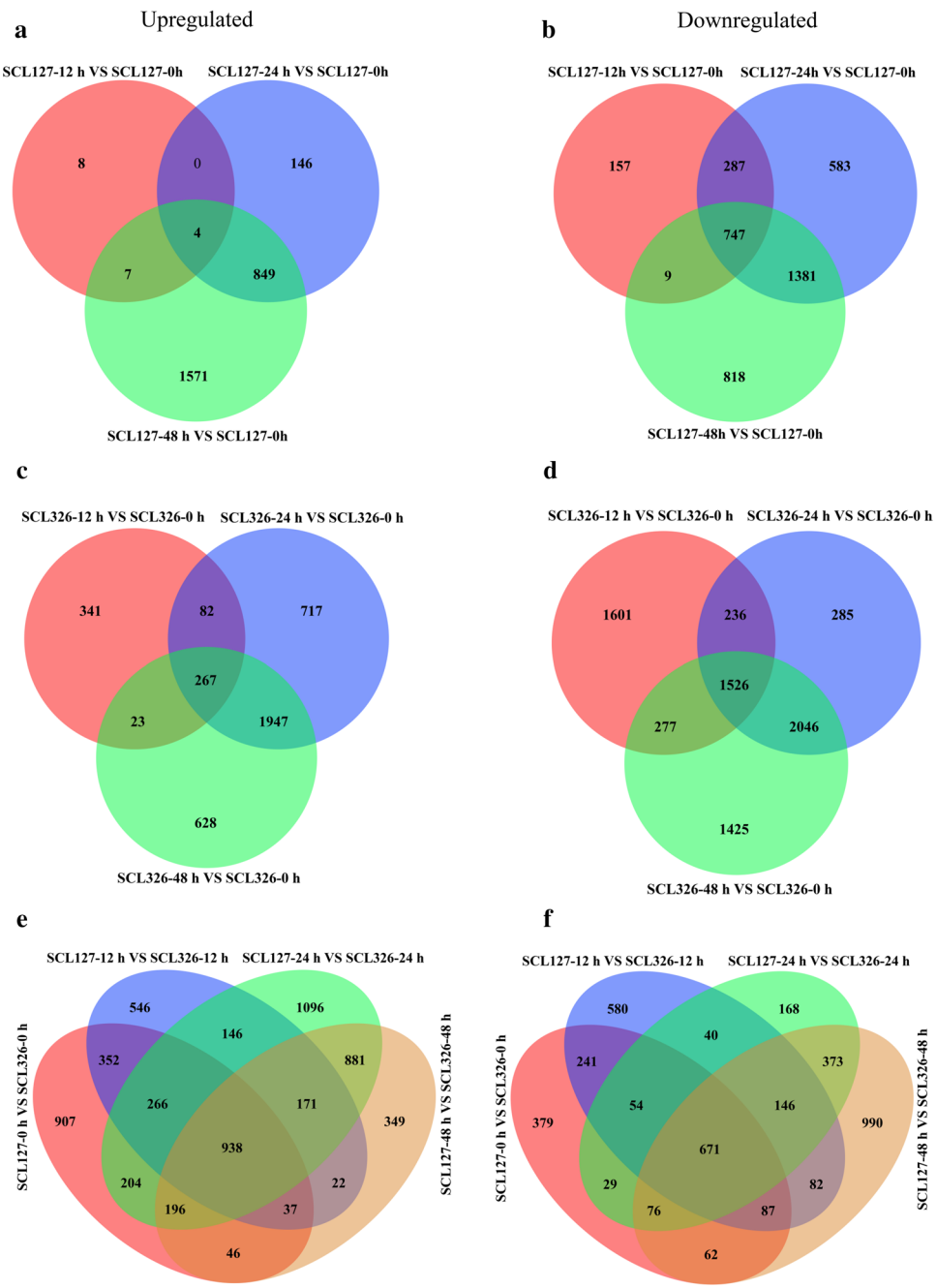

Fig. 2 Venn diagrams of DEGs among different samples. **a**, **c** and **e** Upregulated DEGs among different samples. **b**, **d** and **f** Downregulated DEGs among different samples. DEGs, differentially expressed genes. SCL127-0 h, SCL127-12 h, SCL127-24 h, and SCL127-48 h represent the gene expression levels in the low-germination line (SCL127) at 0, 12, 24, and 48 h of seed germination, respectively. SCL326-0 h, SCL326-12 h, SCL326-24 h, and SCL326-48 h represent the gene expression levels in the high-germination line (SCL326) at 0, 12, 24, and 48 h of seed germination, respectively

Zm00001d003657, *Zm00001d029793*, and *Zm00001d017906* were downregulated with the process of seed germination, whereas *Zm00001d017932* was upregulated in SCL326 (Fig. 3a). The four genes were considered as the priority candidate genes responsible for the seed germination ability in this study.

Intragenic variations affecting the seed germination ability

To reveal the intragenic variations affecting the seed germination ability and identify the favorable haplotypes, we carried out a gene-based association analysis for each of

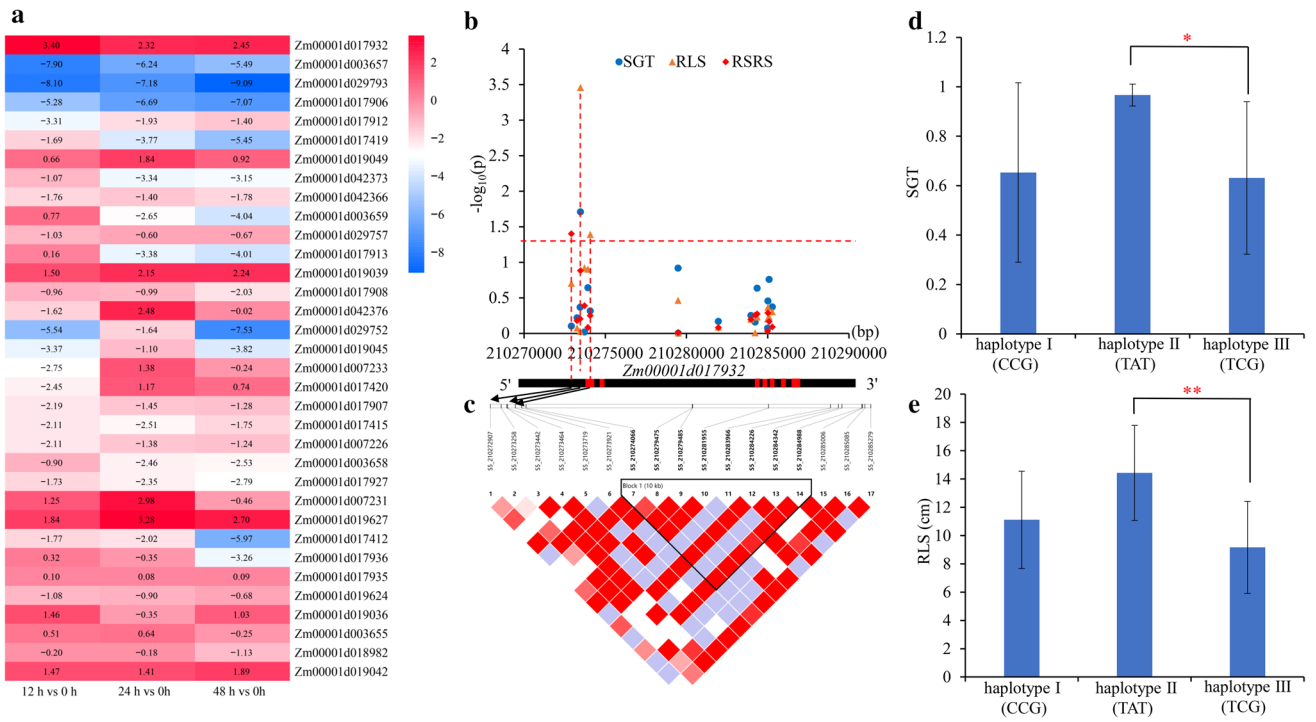

Fig. 3 Combined transcriptome analysis and association study revealing the potential causal gene (*Zm00001d017932*) for seed germination ability in maize. **a** Expression levels of 34 DEGs at different germination stages in SCL326. DEGs, differentially expressed genes. The numbers represent the $\log_2$fold-change of gene expression value between each germination stage and 0 h. Positive and negative numbers represent gene expression was upregulated and downregulated, respectively, in the germination stage relative to 0 h. **b** Significant markers associated with SGT (seed germination rate on 3 d), RLS (root length on 7 d), and RSRS (root–shoot ratio on 7 d). **c** Pairwise LDs (linkage disequilibriums) between the markers. **d** Comparison of SGT (seed germination rate on 3 d) among haplotype I (CCG), haplotype II (TAT), and haplotype III (TCG). *Significant at $P < 0.05$. **e** Comparison of RLS (root length on 7 d) among haplotype I (CCG), haplotype II (TAT), and haplotype III (TCG). **Significant at $P < 0.01$

the priority candidate genes. Among the 68 lines randomly selected from the association panel, a total of 16 (15 SNPs ana 1 InDel), 22 (22 SNPs), 64 (49 SNPs and 15 InDels), 8 (8 SNPs) genetic variations were, respectively, detected in *Zm00001d003657*, *Zm00001d017906*, *Zm00001d017932*, and *Zm00001d029793*. For *Zm00001d017932*, three significant SNPs, S5_210273464 (in the promoter), S5_210272907 (in the promoter), and S5_210274066 (in the first exon), were separately associated with RLS and SGT, RSRS, and RLS, respectively (Fig. 3b). In addition, the eight SNPs, S5_210274066, S5_210279475, S5_210279485, S5_210281955, S5_210283966, S5_210284226, S5_210284342, and S5_210284988, were located within a block, suggesting these markers were closely linked (Fig. 3c). However, no significant variation was detected in the gene body and promoter for the other three genes, *Zm00001d003657*, *Zm00001d017906*, and *Zm00001d029793*. Therefore, we inferred the *Zm00001d017932* as a seed germination ability-associated hub gene in maize.

Based on the three significant SNPs within *Zm00001d017932*, the 68 lines were classed into three major haplotypes. Among these, haplotype II (TAT) had the highest phenotypic values of SGT (0.97) and RLS (14.43 cm), whereas haplotype III (TCG) presented the lowest SGT (0.63) and RLS (9.16 cm) values (Fig. 3d, e). A *t* test showed that significant differences in SGT ($P < 0.05$) and RLS ($P < 0.01$) existed between haplotype II and haplotype III (Fig. 3d, e). Herein, the haplotype with a higher RLS or SGT was designated as a favorable haplotype, and thus, haplotype II (TAT) and haplotype III (TCG) were confirmed as the favorable and unfavorable haplotypes for *Zm00001d017932*, respectively.

Expression patterns of *Zm00001d017932* in different haplotypes

To examine the expression patterns of the *Zm00001d017932* among different haplotypes, we performed the qRT-PCR trials at different seed germination stages (0, 12, 24, and 48 h). The expression abundances of *Zm00001d017932* in the three haplotype II-lines all showed considerable variations at different germination stages relative to 0 h, with the highest expression level at the 12 h of seed germination (Fig. 4a).

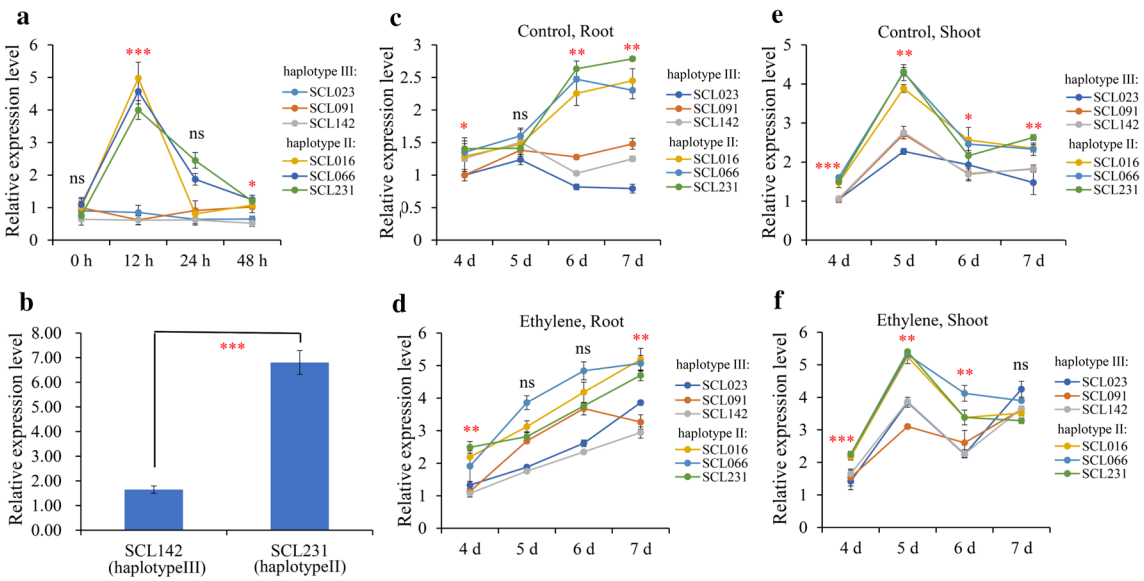

Fig. 4 Expression patterns of *Zm00001d017932*. **a** Expression levels of *Zm00001d017932* among different haplotype lines at different seed germination stages. **b** Expression levels of *Zm00001d017932* in haplotype III (SCL142) and haplotype II (SCL231) at different seed germination stages. **c** Expression levels of *Zm00001d017932* in the roots of contrasting haplotypes at different seed germination stages under control conditions. **d** Expression levels of *Zm00001d017932* in the roots of contrasting haplotypes at different seed germination stages under ethylene treatments. **e** Expression levels of *Zm00001d017932* in the shoots of contrasting haplotypes at different seed germination stages under control conditions. **f** Expression levels of *Zm00001d017932* in the shoots of contrasting haplotypes at different seed germination stages under ethylene treatments. *, **, ***, and ns represent the significance test results between the expression levels of haplotype II and haplotype III at each stage. *, **, and ***Significant at $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively; ns represents not significant

Whereas *Zm00001d017932* in the three haplotype III-lines all displayed slight expression variations at each germination stage, compared with 0 h (Fig. 4a).

To verify whether the *Zm00001d017932* expression was affected by the variations within its promoter, we individually inserted the promoter sequences of haplotype II and haplotype III into the upstream of *GUS* on the vector PCAMBIA3301 and separately transformed them into the tobacco leaves. The relative expression value of *GUS* under the control of haplotype II-promoter was significantly higher (P value < 0.05; approximately 4.13-fold) than that of haplotype III-promoter (Fig. 4b), suggesting these polymorphisms in the promoter of *Zm00001d017932* caused the difference in its expression level between the haplotype II and haplotype III lines.

Zm00001d017932 was annotated as *MADS-transcription factor 26* (*MADS26*), whose homologues were previously reported to influence plant growth and development via ethylene metabolism (Martel et al. 2011; Fujisawa et al. 2013). Thus, we analyzed the expression levels of *MADS26* in different haplotype lines under ethylene treatment. In the roots, the relative expression levels of *MADS26* in haplotype II-lines showed a continuous increase across different germination stages under control conditions (Fig. 4c). However, in the roots of haplotype III-lines, its expression abundance reached a peak on the 5 d and then returned to a lower level following the germination process in the control (Fig. 4c). Under the ethylene treatment, however, *MADS26* presented a continuously upregulated expression in the roots of both haplotypes (Fig. 4d). Generally, the expression patterns of *MADS26* in the shoots of both haplotypes were consistent between the control and ethylene conditions across different germination stages (Fig. 4e, f). Specifically, the expression values of *MADS26* in the shoots of both haplotypes reached the peak on the 5 d in both the control and ethylene conditions, whereas the lowest expression was presented on the 4 d (Fig. 4e, f). Nevertheless, the *MADS26* expression value under the ethylene treatment was higher than that in the control at each of the seed germination stages (Fig. 4e, f). Remarkably, across different germination stages, tissues, and conditions, the *MADS26* expression presented a generally higher level in haplotype II-lines than in haplotype III-lines. This verified these variations in the *MADS26* promoter controlling its expression variations. All these findings also indicated that the *MADS26* expression was induced by ethylene during seed germination. To further verify the ethylene effect on seed germination, we investigated the germination ratio of the haplotype III-lines on the 7 d of ethylene treatment. As a result, the SGS of the haplotype III-lines was significantly ($P < 0.01$) improved to 35.15% under ethylene

treatment in comparison to that of 19.53% under the control (Fig. S6), suggesting that ethylene promoted the germination of maize seeds with low-germination ability.

MADS26 overexpression confers enhanced seed germination ability in *Arabidopsis*

In *Arabidopsis*, the two genes, *FLC* and *AGL67*, controlling seed germination encode MADS-transcription factors. We aligned the protein sequences between FLC and MADS26, and between AGL67 and MADS26. The results showed that the protein homologies of the two pairs both exceed 50.0%, suggesting the *MADS26* gene had the potential to regulate the seed germination in *Arabidopsis*. To verify the role of *MADS26* in seed germination, we performed transgenic assays by overexpressing the *MADS26* CDS of haplotype II (the favorable haplotype) in *Arabidopsis*. Three independent lines (OE1, OE5, and OE6) were selected to investigate the seed germination ability (Fig. 5a). At 48 h of seed germination, the germination rates of three transgenic lines and wide-type Col were all approximately 100% (Fig. S7). However, at 24 h, these transgenic lines displayed significantly enhanced seed germination rates in comparison with the wide-type Col (Fig. 5b, c), with increases of 78.21% (OE1), 89.74% (OE5), and 96.15% (OE6). On the 5 d of seed germination, the root lengths were increased by 0.44-, 0.40-, and 0.48-fold in OE1, OE5, and OE6, respectively, compared with that in the wide-type Col (Fig. 5d, e). These findings demonstrated that *MADS26* improves the seed germination speed and seedling growth in transgenic *Arabidopsis*. Moreover, subcellular localization analyses in tobacco leaf and maize protoplast both showed that the MADS26 protein was targeted to the nucleus (Fig. 5f, g), supporting that *MAD26* acts as a transcription factor.

Discussion

Importance of dissecting the genetic control of maize seed germination ability under optimum conditions

Seed germination ability is a key factor affecting maize production. Previous studies decoded the genetic basis of seed germination ability and identified many QTL under various abiotic stresses (Han et al. 2014; Hu et al. 2016; Li et al. 2018; Zhang et al. 2021). However, only few studies focused on seed germination ability of maize under optimum conditions, and the corresponding genetic foundation is still poorly understood. In this study, we decomposed the seed

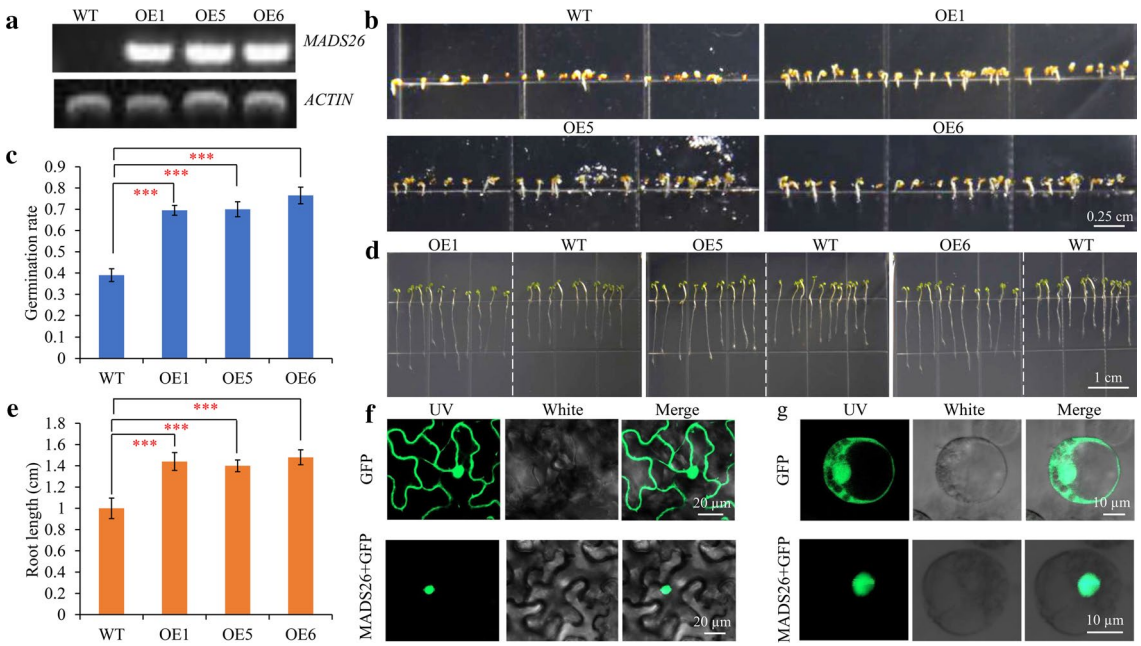

Fig. 5 Subcellular localization of MADS26 protein and overexpression of *MADS26* in *Arabidopsis*. **a** mRNA expression level of *MADS26* in the three transgenic lines of *Arabidopsis*. **b** and **c** Seed germination rate of three transgenic lines at 24 h of germination. ***Significant at $P < 0.001$. **d** and **e** Root length of three transgenic lines on the 5 d of germination. ***Significant at $P < 0.001$. **f** Subcellular localization of MADS26 in tobacco leaf. UV represents ultraviolet. Upper row showed the subcellular localization of GFP, and the bottom row showed the subcellular localization of the MADS26 + GFP fusion protein. **g** Subcellular localization of MAD26 in maize protoplasts. UV represents ultraviolet. Upper row showed the subcellular localization of GFP, and the bottom row showed the subcellular localization of the MADS26 + GFP fusion protein

germination ability of maize into five traits, SGT, SGS, RLS, SLS, and RSRS, under optimum conditions. In the association panel, the phenotypic values of the five traits all showed high variations, with the CVs ranging from 0.36–0.52 (Table S3). This finding suggested that the individuals of the association panel were collected from the germplasms with abundant genetic backgrounds on seed germination ability and suitable for dissecting the genetic architecture of seed germination ability under optimum conditions by using GWAS.

Combined GWAS and transcriptome analysis revealed causal genes involving seed germination ability of maize under optimum conditions

Currently, as an efficient approach, GWAS has been widely used in dissecting the genetic loci controlling complex agronomic traits in crops. Herein, using GWAS, we totally detected 11 significant SNPs associated with seed germination traits of maize under optimum conditions (Fig. 1). Among these SNPs, PZE-102071519, PZE-103101309, and PZE-107033245 were separately situated in the seed germination-associated QTL *qGP1-2*, *qp1GT3-1*, and *qGP2-7* identified in previous studies (Li et al. 2018; Shi et al. 2016). Furthermore, a total of 75 genes were finally found within the LD regions of these significant SNPs (Table S5). The gene model *Zm00001d003657*, involved in the trait SGT, was annotated as a scarecrow protein (Table S5). In *Arabidopsis*, scarecrow-like 15 interacted with histone deacetylase 19 and then repressed the seed maturation (Gao et al. 2014). Moreover, as a key factor, the scarecrow protein was proved to control the root growth in *Arabidopsis* (Koizumi and Gallagher 2013). An SLS-associated gene, *Zm00001d007234*, was annotated as ascorbate peroxidase 2 (Table S5). Overexpressing the ascorbate peroxidase enhanced tobacco seed longevity and germination rates under stress conditions (Lee et al. 2010). The SGT-related gene *Zm00001d017913* encodes a putative leucine-rich repeat receptor-like protein kinase family protein (Table S5), whose homologous gene, *RLK7,* is required for proper germination speed and tolerance to oxidative stress in *Arabidopsis thaliana* (Pitorre et al. 2010). *Zm00001d017932* that was associated with SGT encodes MADS-transcription factor 26 (Table S5). Several studies revealed that the MADS-box participates in the fruit ripening in tomato (Martel et al. 2011; Fujisawa et al. 2013; Shima et al. 2013). The SGS-associated gene, *Zm00001d018984*, was annotated as YlmG homolog protein 1–2 chloroplastic (Table S5). Chen et al. (2018a) reported that a plastid-targeted YlmG protein, EMB1990, was required for chloroplast biogenesis and embryo development in *Arabidopsis*. *Zm00001d019035* encoded a thioredoxin (Table S5), which was implicated in the seed germination in *Medicago truncatula*, wheat, and pea (Lozano et al. 1996; Montrichard et al. 2003; Alkhalfioui et al. 2007). The SGT-associated gene, *Zm00001d042376*, belongs to DUF506 family proteins (Table S5). As a member of DUF506 family, AT3G25240 was demonstrated to negatively regulate the root hair growth in *Arabidopsis* (Ying et al. 2021).

Transcriptome analysis is also an effective method in excavating the DEGs responding to the target traits. However, the single use of GWAS or transcriptome analysis would increase the false positive rate of trait-associated candidate genes (Yao et al. 2020). Thus, a combination of GWAS and transcriptome analysis was considered as a more accurate strategy in detecting the genes controlling agronomic traits. For example, 21 and 7 promising candidate genes involved in *Fusarium* ear rot resistance and seminal root length were, respectively, identified in maize by using this strategy (Guo et al. 2020; Yao et al. 2020). In our study, we totally detected four candidate genes associated with seed germination ability of maize using GWAS and transcriptome analysis. Among the four genes, *Zm00001d017932* that encodes *MADS26* was functionally validated to regulate the seed germination. Since the homologues of *MADS26* were reported to influence plant growth and development via ethylene metabolism (Martel et al. 2011; Fujisawa et al. 2013), we also evaluated the effect of ethylene on seed germination. The results revealed that the ethylene significantly promoted the germination of the seeds with low-germination ratio (Fig. S6). Numerous studies showed that ethylene synergistically or additively acts with GA or CTK to promote seed germination, whereas it antagonizes ABA during the seed germination process (Esashi 2018). Whether the *MADS26* promotes seed germination via above mechanism needs to be confirmed in future studies.

Variations in the *MADS26* promoter affect the seed germination ability

Gene-based association analyses are favorable methods for studying the phenotypic effect caused by variation loci situated within the gene body and promoter region. In this study, the *MADS26*-based association study revealed that three SNPs were significantly correlated with RLS, SGT, and RSRS. According to these significant SNPs, we identified the favorable haplotype (haplotype II: TAT) of *MADS26*. Among them, SNP_210273464 and SNP_210272907 were situated in the promoter of *MADS26*. As well known, the genetic loci variations in promoters probably cause the changes of gene expression abundances. We further validated the effect of the promoter variations on its expression levels by performing qRT-PCR of *MADS26* and transcription activity analysis of its promoter in the contrasting haplotypes. These results both indicated that the promoter of haplotype II had a significantly higher transcription activity than that in haplotype III (Fig. 4b). In addition, previous

studies have demonstrated that favorable haplotypes of genes were closely related to the phenotypic performances of traits (Yu et al. 2019; Li et al. 2020; Zhang et al. 2020). Herein, we transformed the favorable haplotype of CDS (*MADS26*$^{\text{HaplotypeII}}$) into *Arabidopsis* and investigated the seed germination-related traits. The seed germination rate and root length of seedlings in transgenic *Arabidopsis* were significantly higher than those in the wide-type Col, suggesting *MADS26* is a causal factor in regulating seed germination in transgenic *Arabidopsis*. To verify whether the *MADS26* controls the maize seed germination, the favorable haplotype TAT should be given priority for overexpression or knockout in maize. In future, the favorable haplotype TAT can be used for developing functional markers and cultivating high-germination varieties in maize.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00122-022-04065-4.

Author contribution statement YS supervised the project. YS and LM designed the experiments. LM, CW, and YH conducted most of the experiments; WD, ZL, and CZ carried out some experiments. LM and YS wrote the manuscript. LM, YS, GP, and TL edited this draft. All authors have read and approved the manuscript.

Funding This work was supported by the National Key Research and Development Program of China (2021YFF1000303), National Nature Science Foundation of China (32101777 and 31871637), and the Science and Technology Programs of Sichuan Province (2021YJ0476 and 2021JDTD0004).

Data availability All raw data generated of 24 samples used in this study were deposited in Genome Sequence Archive (GSA) in National Genomics Data Center (NGDC) database with the accession number CRA005415.

Declarations

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

Human and animal rights This study does not include human or animal subjects.

References

Alkhalfioui F, Renard M, Vensel WH, Wong J, Tanaka CK, Hurkman WJ, Buchanan BB, Montrichard F (2007) Thioredoxin-linked proteins are reduced during germination of *Medicago truncatula* seeds. Plant Physiol 144(3):1559–1579. https://doi.org/10.1104/pp.107.098103

Anders S, Huber W (2010) Differential expression analysis for sequence count data. Genome Biol 11:R106. https://doi.org/10.1186/gb-2010-11-10-r106

Bassel GW, Lan H, Glaab E, Gibbs DJ, Gerjets T, Krasnogor N, Bonner AJ, Holdsworth MJ, Provart NJ (2011) Genome-wide network model capturing seed germination reveals coordinated regulation of plant cellular phase transitions. Proc Natl Acad Sci USA 108(23):9709–9714. https://doi.org/10.1073/pnas.1100958108

Bewley JD, Bradford KJ, Hilhorst H (2012) Seeds: physiology of development, germination and dormancy. Springer Science & Business Media, New York

Cao X, Becker LC, Becker DM, Starmer JD, Province MA (2010) Avoiding the high bonferroni penalty in genome-wide association studies. Genet Epidemiol 34(1):100–105. https://doi.org/10.1002/gepi.20430

Chen H, Li S, Li L, Hu H, Zhao J (2018a) *Arabidopsis EMB1990* encoding a plastid-targeted YlmG protein is required for chloroplast biogenesis and embryo development. Front Plant Sci 9:181. https://doi.org/10.3389/fpls.2018.00181

Chen S, Zhou Y, Chen Y, Gu J (2018b) fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34(17):i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chiang GCK, Barua D, Kramer EM, Amasino RM, Donohue K (2009) Major flowering time gene, *FLOWERING LOCUS C*, regulates seed germination in *Arabidopsis thaliana*. Proc Natl Acad Sci USA 106(28):11661–11666. https://doi.org/10.1073/pnas.0901367106

Clough SJ, Bent AF (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. Plant J 16(6):735–743. https://doi.org/10.1046/j.1365-313x.1998.00343.x

Devaiah SP, Pan X, Hong Y, Roth M, Welti R, Wang X (2007) Enhancing seed quality and viability by suppressing phospholipase D in *Arabidopsis*. Plant J 50(6):950–957. https://doi.org/10.1111/j.1365-313X.2007.03103.x

Esashi Y (2018) Ethylene and seed germination. The plant hormone ethylene. CRC press: 133–157

Fujisawa M, Nakano T, Shima Y, Ito Y (2013) A large-scale identification of direct targets of the tomato MADS box transcription factor RIPENING INHIBITOR reveals the regulation of fruit ripening. Plant Cell 25(2):371–386. https://doi.org/10.1105/tpc.112.108118

Gao MJ, Li X, HuangJ GGM, Gjetvaj B, Lindsay DL, Wei S, Coutu C, Chen Z, Wan XC, Hannoufa A, Lydiate DJ, Gruber MY, Chen ZJ, Hegedus DD (2014) SCARECROW-LIKE15 interacts with HISTONE DEACETYLASE19 and is essential for repressing the seed maturation programme. Nat Commun 6:7243. https://doi.org/10.1038/ncomms8243

Guo T, Yang J, Li D, Sun K, Luo L, Xiao W, Wang J, Liu Y, Wang S, Wang H, Chen Z (2019) Integrating GWAS, QTL, mapping and RNA-seq to identify candidate genes for seed vigor in rice (*Oryza sativa* L.). Mol Breeding 39(6):1–16. https://doi.org/10.1007/s11032-019-0993-4

Guo J, Li C, Zhang X, Li Y, Zhang D, Shi Y, Song Y, Li Y, Yang D, Wang T (2020) Transcriptome and GWAS analyses reveal candidate gene for seminal root length of maize seedlings under drought stress. Plant Sci 292:110380. https://doi.org/10.1016/j.plantsci.2019.110380

Han Z, Ku L, Zhang Z, Zhang J, Guo S, Liu H, Zhao R, Ren Z, Zhang L, Su H, Dong L, Chen Y (2014) QTLs for seed vigor-related traits identified in maize seeds germinated under artificial aging conditions. PLoS ONE 9(3):e92535. https://doi.org/10.1371/journal.pone.0092535

Hartweck LM, Scott CL, Olszewski NE (2002) Two O-linked *N*-acetylglucosamine transferase genes of *Arabidopsis thaliana* L. Heynh. have overlapping functions necessary for gamete and seed development. Genetics 161(3):1279–1291. https://doi.org/10.1093/genetics/161.3.1279

Henschel K, Kofuji R, Hasebe M, Saedler H, Münster T, Theißen G (2002) Two ancient classes of MIKC-type MADS-box genes are present in the moss *Physcomitrella* patens. Mol Biol Evol

19(6):801–814. https://doi.org/10.1093/oxfordjournals.molbev.a004137

Hu S, Lübberstedt T, Zhao G, Lee M (2016) QTL mapping of low-temperature germination ability in the maize IBM Syn4 RIL population. PLoS ONE 11(3):e0152795. https://doi.org/10.1371/journal.pone.0152795

Hu S, Sanchez DL, Wang C, Lipka AE, Yin Y, Gardner CAC, Lübberstedt T (2017) Brassinosteroid and gibberellin control of seedling traits in maize (*Zea mays* L.). Plant Sci 263:132–141. https://doi.org/10.1016/j.plantsci.2017.07.011

Jiao Y, Peluso P, Shi J, Liang T, Stitzer MC, Wang B, Campbell MS, Stein JC, Wei X, Chin CS, Guill K, Regulski M, Kumari S, Olson A, Gent J, Schneider KL, Wolfgruber TK, May MR, Springer NM, Antoniou E, Richard McCombie W, Presting GG, McMullen M, Ross-Ibarra J, Kelly Dawe R, Hastie A, Rank DR, Ware D (2017) Improved maize reference genome with single-molecule technologies. Nature 546(7659):524–527. https://doi.org/10.1038/nature22971

Johnson RC, Nelson GW, Troyer JL, Lautenberger JA, Kessing BD, Winkler CASJ, O'Brien (2010) Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics 11(1):1–6. https://doi.org/10.1186/1471-2164-11-724

Katsuya-Gaviria K, Caro E, Carrillo-Barral N, Iglesias-Fernández R (2020) Reactive oxygen species (ROS) and nucleic acid modifications during seed dormancy. Plants 9(6):679. https://doi.org/10.3390/plants9060679

Kim YO, Pan SO, Jung CH, Kang H (2007) A zinc finger-containing glycine-rich RNA-binding protein, atRZ-1a, has a negative impact on seed germination and seedling growth of *Arabidopsis thaliana* under salt or drought stress conditions. Plant Cell Physiol 48(8):1170–1181. https://doi.org/10.1093/pcp/pcm087

Koizumi K, Gallagher KL (2013) Identification of SHRUBBY, a SHORT-ROOT and SCARECROW interacting protein that controls root growth and radial patterning. Development 140(6):1292–1300. https://doi.org/10.1242/dev.090761

Kucera B, Cohn MA, Leubner-Metzger G (2005) Plant hormone interactions during seed dormancy release and germination. Seed Sci Res 15(4):281–307. https://doi.org/10.1079/ssr2005218

Landjeva S, Lohwasser U, Börner A (2010) Genetic mapping within the wheat D genome reveals QTL for germination, seed vigour and longevity, and early seedling growth. Euphytica 171(1):129–143. https://doi.org/10.1007/s10681-009-0016-3

Latha M, Dolui AK, Vijayaraj P (2021) Proteoform of *Arabidopsis* seed storage protein identified by functional proteomics approach exhibits acyl hydrolase activity during germination. Int J Biol Macromol 172:452–463. https://doi.org/10.1016/j.ijbiomac.2021.01.074

Lee YP, Baek KH, Lee HS, Kwak SS, Bang JW, Kwon SY (2010) Tobacco seeds simultaneously over-expressing Cu/Zn-superoxide dismutase and ascorbate peroxidase display enhanced seed longevity and germination rates under stress conditions. J Exp Bot 61(9):2499–2506. https://doi.org/10.1093/jxb/erq085

Li X, Wang G, Fu J, Li L, Jia G, Ren L, Lubberstedt T, Wang G, Wang J, Gu R (2018) QTL mapping in three connected populations reveals a set of consensus genomic regions for low temperature germination ability in *Zea mays* L. Frontiers Plant Sci 9:65. https://doi.org/10.3389/fpls.2018.00065

Li Z, Liu P, Zhang X, Zhang Y, Ma L, Liu M, Guan Z, Zhang Y, Li P, Zou C, He Y, Gao S, Pan G, Shen Y (2020) Genome-wide association studies and QTL mapping uncover the genetic architecture of ear tip-barrenness in maize. Physiol Plantarum 170(1):27–39. https://doi.org/10.1111/ppl.13087

Lipka AE, Tian F, Wang Q, Peifer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Liu L, Lai Y, Cheng J, Wang L, Du W, Wang Z, Zhang H (2014) Dynamic quantitative trait locus analysis of seed vigor at three maturity stages in rice. PLoS ONE 9(12):e115732. https://doi.org/10.1371/journal.pone.0115732

Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet 12:e1005767. https://doi.org/10.1371/journal.pgen.1005767

Lozano RM, Wong JH, Yee BC, Peters A, Kobrehel K, Buchanan BB (1996) New evidence for a role for thioredoxin h in germination and seedling development. Planta 200:100–106. https://doi.org/10.1007/bf00196655

Ma L, Liu M, Yan Y, Qing C, Zhang X, Zhang Y, Long Y, Wang L, Pan L, Zou C, Li Z, Wang Y, Peng H, Pan G, Jiang Z, Shen Y (2018) Genetic dissection of maize embryonic callus regenerative capacity using multi-locus genome-wide association studies. Front Plant Sci 9:561. https://doi.org/10.3389/fpls.2018.00561

Ma L, Qing C, Frei U, Shen Y, Lübberstedt T (2020) Association mapping for root system architecture traits under two nitrogen conditions in germplasm enhancement of maize doubled haploid lines. Crop J 8(2):213–226. https://doi.org/10.1016/j.cj.2019.11.004

Ma L, Qing C, Zhang M, Zou C, Pan G, Shen Y (2021) GWAS with a PCA uncovers candidate genes for accumulations of microelements in maize seedlings. Physiol Plantarum 172(4):2170–2180. https://doi.org/10.1111/ppl.13466

Martel C, Vrebalov J, Tafelmeyer P, Giovannoni JJ (2011) The tomato MADS-box transcription factor RIPENING INHIBITOR interacts with promoters involved in numerous ripening processes in a COLORLESS NONRIPENING-dependent manner. Plant Physiol 157(3):1568–1579. https://doi.org/10.1104/pp.111.181107

Masiero S, Colombo L, Grini PE, Schnittger A, Kater MM (2011) The emerging importance of type I MADS box transcription factors for plant reproduction. Plant Cell 23(3):865–872. https://doi.org/10.1105/tpc.110.081737

Montrichard F, Renard M, Alkhalfioui F, Duval FD, Macherel D (2003) Identification and differential expression of two thioredoxin *h* isoforms in germinating seeds from pea. Plant Physiol 132(3):1707–1715. https://doi.org/10.1104/pp.102.019562

Oge L, Bourdais G, Bove J, Collet B, Godin B, Granier F, Boutin J-P, Job D, JullienM GP (2008) Protein repair Lisoaspartyl methyltransferase1 is involved in both seed longevity and germination vigor in *Arabidopsis*. Plant Cell 20(11):3022–3037. https://doi.org/10.1105/tpc.108.058479

Pace J, Gardner C, Romay C, Ganapathysubramanian B, Lübberstedt T (2015) Genome-wide association analysis of seedling root development in maize (*Zea mays* L.). BMC Genomics 16(1):1–12. https://doi.org/10.1186/s12864-015-1226-9

Pitorre D, Llauro C, Jobet E, Guilleminot J, Brizard JP, Delseny M, Lasserre E (2010) RLK7, a leucine-rich repeat receptor-like kinase, is required for proper germination speed and tolerance to oxidative stress in *Arabidopsis thaliana*. Planta 232(6):1339–1353. https://doi.org/10.1007/s00425-010-1260-4

Ren X, Pan Z, Zhao H, Zhao J, Cai M, Li J, Zhang Z, Qiu F (2017) EMPTY PERICARP11 serves as a factor for splicing of mitochondrial nad1 intron and is required to ensure proper seed development in maize. J Exp Bot 68(16):4571–4581. https://doi.org/10.1093/jxb/erx212

Rochat C, Boutin JP (1992) Temporary storage compounds and sucrose-starch metabolism in seed coats during pea seed development (*Pisum sativum*). Physiol Plantarum 85(4):567–572. https://doi.org/10.1111/j.1399-3054.1992.tb04756.x

Shi Y, Li G, Tian Z, Wang Z, Wang X, Zhu Y, Chen Y, Guo S, Qi J, Zhang X, Ku L (2016) Genetic dissection of seed vigour traits in maize (*Zea mays* L.) under low-temperature conditions. J Genet 95(4):1017–1022. https://doi.org/10.1007/s12041-016-0714-2

Shima Y, Kitagawa M, Fujisawa M, Nakano T, Kato H, Kimbara J, Kasumi T, Ito Y (2013) Tomato FRUITFULL homologues act in fruit ripening via forming MADS-box transcription factor complexes with RIN. Plant Mol Biol 82(4–5):427–438. https://doi.org/10.1007/s11103-013-0071-y

Shin JH, Kim SR, An G (2009) Rice aldehyde dehydrogenase7 is needed for seed maturation and viability. Plant Physiol 149(2):905–915. https://doi.org/10.1104/pp.108.130716

Suzuki Y, Yasui T, Matsukura U, Terao J (1996) Oxidative stability of bran lipids from rice variety [*Oryza sativa* (L.)] lacking lipoxygenase-3 in seeds. J Agric Food Chem 44(11):3479–3483. https://doi.org/10.1021/jf9600465

Tian Y, Dong Q, Ji Z, Chi F, Cong P, Zhou Z (2015) Genome-wide identification and analysis of the MADS-box gene family in apple. Gene 555(2):277–290. https://doi.org/10.1016/j.gene.2014.11.018

Weber H, Borisjuk L, Wobus U (1997) Sugar import and metabolism during seed development. Trends Plant Sci 2(5):169–174. https://doi.org/10.1016/s1360-1385(97)85222-3

Yao L, Li Y, Ma C, Tong L, Du F, Xu M (2020) Combined genome-wide association study and transcriptome analysis reveal candidate genes for resistance to *Fusarium* ear rot in maize. J Integr Plant Biol 62(10):1535–1551. https://doi.org/10.1111/jipb.12911

Ying S, Blancaflor E B, Liao F, Scheible WR (2021) A phosphorus-limitation induced, functionally conserved DUF506 protein is a repressor of root hair lonegation in *Arabidopsis thaliana*. bioRxiv: https://doi.org/10.1101/2021.07.09.451837

Yoo SD, Cho YH, Sheen J (2007) *Arabidopsis* mesophyll protoplasts: a versatile cell system for transient gene expression analysis. Nat Protoc 2(7):1565–1572. https://doi.org/10.1038/nprot.2007.199

Yu F, Liang K, Fang T, Zhao H, Han X, Cai M, Qiu F (2019) A group VII ethylene response factor gene, *ZmEREB180*, coordinates waterlogging tolerance in maize seedlings. Plant Biotechnol J 17(12):2286–2298. https://doi.org/10.1111/pbi.13140

Yuan W, Flowers JM, Sahraie DJ, Ehrenreich IM, Purugganan MD (2016) Extreme QTL mapping of germination speed in *Arabidopsis thaliana*. Mol Ecol 25(17):4177–4196. https://doi.org/10.1111/mec.13768

Zhang X, Zhang H, Li L, Lan H, Ren Z, Liu D, Wu L, Jaqueth J, Li B, Pan G, Gao S (2016) Characterizing the population structure and genetic diversity of maize breeding germplasm in Southwest China using genome-wide SNP markers. BMC Genomics 17(1):1–16. https://doi.org/10.1186/s12864-016-3041-3

Zhang Y, Hu Y, Guan Z, Liu P, He Y, Zou C, Li P, Gao S, Peng H, Yang C, Pan G, Shen Y, Ma L (2020) Combined linkage mapping and association analysis reveals genetic control of maize kernel moisture content. Physiol Plantarum 170(4):508–518. https://doi.org/10.1111/ppl.13180

Zhang Y, Liu P, Wang C, Zhang N, Zhu Y, Zou C, Yuan G, Yang C, Gao S, Pan G, Ma L, Shen Y (2021) Genome-wide association study uncovers new genetic loci and candidate genes underlying seed chilling-germination in maize. PeerJ 9:e11707. https://doi.org/10.7717/peerj.11707