New-Phytologist- Supporting Information for

Myb10-D confers *PHS-3D* resistance to pre-harvest sprouting by regulating *NCED* in ABA biosynthesis pathway of wheat

Jing Lang, Yuxin Fu, Yong Zhou, Mengping Cheng, Min Deng, Maolian Li, Tingting Zhu, Jian Yang, Xiaojiang Guo, Lixuan Gui, Linchuan Li, Zhongxu Chen, Yingjin Yi, Lianquan Zhang, Ming Hao, Lin Huang, Chao Tan, Guoyue Chen, Qiantao Jiang, Pengfei Qi, Zhien Pu, Jian Ma, Zehou Liu, Yujiao Liu, Ming-Cheng Luo, Yuming Wei, Youliang Zheng, Yongrui Wu, Dengcai Liu and JiRui Wang

Article acceptance date: 16 February 2021

The following Supporting Information is available for this article:

Supplementary Methods S1-S18

Supplementary Tables S1-S9

Supplementary Figures S1-S6

Supplementary Methods

S1 DNA extraction and genotyping

Young leaf tissues were used to isolate genomic DNA for PCR and genomic resequencing (Murray & Thompson, 1980). Genomic DNA of SHW-L1 and AS60 were sent to BerryGenomics (Beijing) for resequencing (Table S3).

The sequences of primers for KASP markers, STS markers are listed in Table S1. All primers were generated using IDT (https://sg.idtdna.com/pages), NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi?LINK_LOC=Blast Home) online tools, and DNAMAN.

For the Kompetitive allele-specific PCR (KASP) genotyping of fine mapping lines, a PCR system, Model CFX96 (BioRad, USA), was used. Microplates with 96-wells were used with a 10µl total reaction volume in each well. The PCR reactions used LGC (KBS-1016-002) and consisted of: 5µl 2×KASP Master mix, 2.52µl H2O, 1.4µl primer mix, 0.08µl MgCl2 (50mM), 1µl DNA template (ca. 100 ng/µl). PCR amplification was conducted using a touchdown program that started at 94°C for 15min. Followed by 10 cycles of 94°C for 20s, 65-55°C (drop 1 °C per cycle) for 60s. And followed by 38 cycles of 94°C for 20s, 55°C for 60 s. With a final extension at 37°C for 60s to read the fluorescence signal value.

For STS-genotyping of wheat germplasms, the PCR reactions Vazyme (P505) was used that consisted of 10.4 µl H2O, 15µl 2×Phanta Max Buffer, 1.2 µl FP (10µM), 1.2 µl RP (10µM), 0.6 µl Phanta Max Super-Fidelity DNA Polymerase, 0.6 µl dNTP Mix (10 mM each), 1 µL DNA template (ca. 100 ng/µl). PCR amplification was conducted using a program that started at 95°C for 3min. Followed by 35 cycles (95°C for 20s, 50-65°C for 30s, 72°C for 60s), and 72°C for 10min. The desired fragments were purified by FastPure® Gel DNA Extraction Mini Kit (Vazyme, http://www.vazyme.com/Home.html), and ligated with pEASY®-Blunt Cloning Kit (TransGen http://www.transgenbiotech.com/). Three positive clones of STS products for each marker of CS and AK58 were sequenced by Bioengineering (China Shanghai) Co., Ltd., with ABI 3730 sequencing, and then the primes were used for germplasms genotyping.

S2 Fine mapping of *PHS-3D* (*QPHS.sicau-3D*)

GP 0.8 and GP 0.5 were used as the signal of seed dormancy breaking in wheat research and production, respectively (Xiao et al, 2002). In total, 3,300 $F_{3:4}$ lines were randomly selected from 7,500 $F_{3:4}$ lines for both PHS resistance evaluation and KASP genotyping. For classification of these lines, the germination rate of less than 0.50 was defined as a resistant material, 910 of them showed PHS resistance. Germination rate greater than 0.80 is defined as susceptive material, 836 of them showed PHS susceptive. And 1,476 lines showed moderate PHS resistance. Based on KASP genotyping, 18 types of recombinants were obtained with abundant recombinants, which from the 3,222 $F_{3:4}$ lines.

S3 Optical map construction

Ultra-high molecular weight (uHMW) DNA was isolated from young leaves of CM32 using the Plant DNA Isolation Kit (Bionano Genomics, San Diego, CA). The uHMW DNA molecules were labeled with DLE-1 enzyme (Bionano Genomics, San Diego, CA), and were then stained according to the instructions of Bionano Prep™ Direct Label and Stain (DLS) Kit (Bionano Genomics, San Diego, CA). The labeled DNA was then loaded onto the Saphyr platform (Bionano Genomics, San Diego, CA), and the raw molecules were collected. A consensus optical map was *de novo* assembled with the Assembler tool of the Bionano Solve v3.4 package using significance cutoffs of $P < 1 \times 10^{-9}$ to generate draft consensus contigs, $P < 1 \times 10^{-10}$ for draft consensus contig extension, and $P < 1 \times 10^{-15}$ for the final merging of the draft consensus contigs; and a recipe of "nonhaplotype", "noES", and "noCut" was chosen.

The alignment of the optical map to the sequences was performed using the RefAligner tool of the Bionano Solve v3.4 package (Bionano Genomics, San Diego, CA), with a cutoff of $P<1 \times 10^{-10}$. The result was visualized in the Bionano Access v1.4 software.

S4 Gene cloning and sequencing

Primers used for vectors pENTR/D-TOPO (Invitrogen K240020) and Gateway-destination-1300 construction, and positive clone identification are listed in Table S1. For *Myb10-D* cloning, the Vazyme (P505) PCR reactions were used. PCR amplification was conducted using a touchdown program that started at 95°C for 3min, followed by

35 cycles of 95°C for 15 s, 62°C for 15s, and 72°C for 2min, with a final extension at 72°C for 10min. The PCR products were analyzed in a 2.0% agarose gel. The PCR fragments were purified using the Tiangen Gel DNA Recovery kit (http://www.tiangen.com/en/?product1/t1/4.html) and then ligated with pEASY®-Blunt Cloning Kit following the manufacturer's instructions. Three positive clones were sequenced by TSINGKE (China Chengdu) Co., Ltd. with ABI 3730 sequencing.

S5 RNA preparation and RNA sequencing

The parents of RILs SHW-L1 and CM32, transgenic T2, T3 lines (OE), and wild control (WT) were grown in a growth chamber for the RNA-seq experiment at 20 °C for 16h of light and 16 °C for 8h of darkness. Whole spikes were collected at 5, 10, 15, 20, 25, and 30 days post-anthesis (DPA). Three to four seeds in the central parts of each spike were flashed frozen in liquid nitrogen and stored at −80°C for RNA extracting.

Total RNA was extracted from wheat 5, 10, 15, 20, 25, and 30DPA seeds using the RNAplant plus Reagent (DP437) (Tiangen) according to the manufacturer's instructions and sent to BGI Co Ltd (www.genomics.cn) for RNA-seq. RNA quantity and quality were assessed using NanoDrop 2000 spectrophotometer (Thermo Scientific, USA) and checked for integrity on an Agilent 2100 bioanalyzer (Agilent Technologies, USA) by denaturing agarose gel electrophoresis with ethidium bromide staining.

Equal molar amounts of the libraries f were constructed and sequenced by BerryGenomics (Beijing) using the Illumina HiSeq-2000 and HiSeq X Ten platform (Illumina, USA). Two replicated sets of RNA-seq data were combined for analysis to increase the sequencing depth (Table S2). The original fluorescence image files were transformed into short reads (raw data) by base calling, and these reads were recorded in FASTQ format. Data processing were following the instructions of BerryGenomics (Beijing).

S6 NGS data analysis: reads mapping, assembly, and functional annotation

The genomic variance between CS (552.0-592.0 Mb) and AK58 (563.0-603.0 Mb) was carried in D-Genies (http://dgenies.toulouse.inra.fr/) (Cabanettes & Klopp, 2018). Valid RNA-seq and resequencing data were sequenced by BerryGenomics (Beijing) following their instructions, which were mapped to the genomic reference sequence

of CS 2.0 (IWGSC, ftp://ftp.ensemblgenomes.org/pub/plants/current/fasta/triticum_aestivum/dna/) by Burrows-Wheeler Aligner (BWA) version 0.7.17 to get the original mapping results. If one or paired reads were mapped to multiple positions, the strategy adopted by BWA was to choose the most likely placement. Then, SAMtools (Li *et al.*, 2009) and GATK (DePristo *et al.*, 2011) were used to sort BAM files and do duplicate marking, local realignment, and base quality recalibration to generate the final BAM file for computation of the sequence coverage and depth. The mapping step was challenging due to mismatches, including true mutation and sequencing error, and duplicates resulted from PCR amplification. These duplicate reads were uninformative and shouldn't be considered as evidence for variants. Picard was employed to mark these duplicates so that GATK would ignore them in the following analysis. BCFtools (version 4.0) was used to do the variant calling and identify SNP, and InDels. Genomic sequence assembly of the processed reads of SHW-L1 and AS60 were carried out using BCFtools (bcftools consensus -f split_genome.fa $sample_name.vcf.gz –o; output_$sample_name.fa). The Kallisto (version 0.43.0) was used to assemble transcripts and identify differentially expressed genes. Gene transcript levels were estimated using TPM (Transcripts per Million). The differentially expressed genes were BLAST searched against the CS reference sequence. Genes in major metabolic pathways were identified by enrichment factor analyses. Gene annotation was obtained from IWGSC (http://www.wheatgenome.org/ Tools-and-Resources/IWGSC-RefSeq-Annotation-v2.0-call-for-contributions) Specific unigenes were used as query sequences to search the annotation databases, including the NCBI non-redundant protein sequences database (NR) (http://www.ncbi.nlm.nih.gov/), Kyoto Encyclopedia of Genes and Genome (KEGG) pathways database (http://www.genome.jp/kegg/), and based on sequence homology to entries in the Gene Ontology (GO) database (http://www.geneontol ogy.org/). We classified genes according to KEGG classifications and conducted enrichment analyses using a cluster profile.

Transposable Elements were characterized by TREP (Transposable Elements Platform, https://botserv2.uzh.ch/kelldata/trep-db/blast/blastTREP.html).

S7 Quantitative PCR with reverse transcription (RT-PCR)

Total RNA was extracted from wheat 5, 10, 15, 20, 25, and 30DPA seeds using the RNAplant plus Reagent (DP437) (Tiangen) according to the manufacturer's instructions. cDNA was synthesized by a kit (TaKaRa, RR047A). Quantitative real-time PCR was performed on the BioRad CFX96 instrument using SYBR Green Q711 (Vazyme). The reaction volume consisted of 10µl 2×ChamQ Universal SYBR qPCR Master Mix, 0.4 µl FP (10µM), 0.4 µl RP (10µM), 1.2µl cDNA, and 8 µl H_2O. RT-PCR in a condition of 95°C for 30 s, followed by 40 cycles of 95 °C for 10 s, 60 °C for 30 s, 72 °C for 30s. For all samples, three reference genes *TaActin*, *Ta.14126.1*, and *Ta.7894.3.A1,* were used as internal control genes (Long *et al.*, 2010). The date was read by IQ5 software. The ultimate levels of templates tested by internal control genes were statistically operated by geometric mean. In the end, the relative expression quantity of each sample was calculated using $E^{-\Delta\Delta Ct}$ the method (Pfaffl, 2001). Thus, relative expression quantities of RILs were used for followed eQTL analysis.

S8 Gene transformation

The pENTR-Myb10-D plasmid was constructed using the Gateway pENTR™/D-TOPO™ (invitrogen K240020, USA) after its genomic sequence (2086bp) fragment was amplified from SHW-L1 using primers TaMyb10-D-F7.1 and TaMyb10-D-R7. LR recombination Gateway® reaction (ThermoFisher Scientific, USA) were subsequently performed between the entry clone pENTR-Myb10-D and the correlative destination vectors pCAMBIA1300.

We also use a binary vector (pUbi-CAMBIA3301) containing the Bar selection gene to transform full-length cDNA of *Myb10-D* according to pUBI:: *Myb10-D* construct.

The pUBI:: *Myb10-D* plasmid was constructed using the Homologous Recombination Monad (MC40101) after its cDNA sequence (798bp) fragment was amplified from SHW-L1 using primers TaMyb10-D-cDNA-F and TaMyb10-D-cDNA-R.

The pCAMBIA1300-Myb10-D and pUBI:: *Myb10-D* were transformed into the Agrobacterium tumefaciens strain EHA105 and introduced into wheat cultivar Fielder by Bangdi Biotec Company (Jinan, China) and GEB Biotec Company (Beijing, China) (Ishida *et al.*, 2015). And the primers used in this study are listed in Table S1.

Twelve independent plants regenerated from the transformation experiment were grown in growth chambers. Transgene copy number analyses were carried out in T1

and T2 generations with the primer ECA and HYG-3 (Mieog *et al.*, 2013). In total, 104 and 26 T1 progenies from 2 pCAMBIA1300 and five pUbi-CAMBIA3301 positive T0 plants were used in the progeny test. Because the positive plants of both vectors had similar phenotypes, we select pCAMBIA1300 for the following analysis. T1 plants were allowed to self-pollinate, and the resulting progeny T2 and T3 were assessed for phenotyping (grain color, seed germination, seed coat permeability, RNA-seq, ABA, and flavonoids concentration test).

S9 Endogenous ABA profiling

Myb10-D transgenic T2 lines (OE) and wild control (WT) were grown in a growth chamber for plant metabolomics profiling at 20°C for 16h of light and at 16°C for 8h of darkness. Whole spikes were collected at DPA20. One gram seeds in the central parts of peaks were flashed frozen in liquid nitrogen and stored at −80°C for ABA profiling. Plant materials were ground into a powder and extracted with methanol/water/formic acid (15:4:1, V/V/V). The combined extracts were evaporated to dryness under nitrogen gas stream, reconstituted in 80% methanol (V/V), and filtrated (PTFE, 0.22 μm; Anpel) before LC-MS/MS analysis. Three replicates of each assay were performed. ABA contents were detected by MetWare (http://www.metware.cn/) based on the AB Sciex QTRAP 6500 LC-MS/MS platform (Floková *et al.*, 2014).

S10 Flavonoids extraction and profiling (multiple reaction monitoring, MRM)

Myb10-D transgenic T2 lines (OE) and wild control (WT) were grown in a growth chamber for plant metabolomics profiling at 20°C for 16h of light and at 16°C for 8h of darkness. Thirty seed (at least 1 gram) from 5 independent plants (only seeds from spikelets at the central spike from the main spike that were first following were used) were combined for each of three replications. The freeze-dried grain was crushed using a mixer mill (MM 400, Retsch) with a zirconia bead for 1.5 min at 30 Hz. 100mg powder was weighted and extracted overnight at 4°C with 1.0 ml 70% aqueous methanol. Following centrifugation at 10000g for 10 min, the extracts were absorbed (CNWBOND Carbon-GCB SPE Cartridge, 250mg, 3mL; ANPEL, Shanghai, China, www.anpel.com.cn/cnw) and filtrated (SCAA-104, 0.22μm pore size; ANPEL, Shanghai, China, http://www.anpel.com.cn/) before LCMS analysis. Flavonoids contents were

detected by MetWare (http://www.metware.cn/) based on HPLC and ESI-Q TRAP-MS/MS platform (Schultz *et al.*, 2013).

S11 Subcellular localization of *Myb10-D*

The amino acid sequences of *Myb10-D* were used to predict the nuclear localization signal (NLS) by online tools wolf-psort (https://wolfpsort.hgc.jp) or SoftBerry (http://linux1.softberry.com/berry.phtml?topic=protcomppl&group=programs&subgroup=proloc). To determine the subcellular localization of *Myb10-D*, 35S::GFP-Myb10-D expression vector was constructed by fusing the coding sequence of *Myb10-D* from SHW-L1 to a GFP coding sequence. The coding regions were amplified by specific primer (F: ATGGGGAGGAAGCCATGCTGCGCCA; R: CTAGCAAAGCCACGCCAACTCCAGG). The C-terminal of *Myb10-D* was fused to a reporter gene encoding the enhanced GFP (eGFP) to get 35S::*Myb10-D*:GFP, which was inserted into pCAMBIA1300 plasmid driven by 35S promoter by primers (Myb10-D-KpnI-F: CGGGGTACCATGGGGAGGAAGCCATGCTGCGC; Myb10-D-XbaI-R: CTAGTCTAGAGCAAAGCCACGCCAACTCCAGG). Agrobacterium-mediated transient transformation of *Nicotiana benthamiana* plants was conducted.

The *N. Benthamiana* plants were grown in a greenhouse at 25 °C for 16h of light and 20 °C for 8h of darkness. For subcellular localization analysis, CDS sequence of *Myb10-D* was cloned into between KpnI and XbaI sites of the pCAMBIA1301 vector to generate pCAMBIA1301-*Myb10-D* (p35S:*Myb10-D*:GFP).
Vectors were transiently transfected into *N. Benthamiana* leaves (Liu *et al.*, 2010). Fluorescence was examined under a confocal microscopy (FV1000, Olympus, Japan) at 2 days after transformation.

S12 RNA *in situ* hybridization

The tissue expression pattern of *Myb10-D* in the cultivar Fielder (WT) and transgenic overexpression lines (OE) were assayed by RNA in situ hybridization. Tissue processing and in situ hybridization experiments were carried out according to Janes (Janes *et al.*, 2010). The seeds were fixed in 4% paraformaldehyde solution (sigma) with 0.1% Triton X-100 (Sigma) and 0.1% Tween-20 in PBS (Takara, cat #900) for 16h. After dehydration using graded ethanol and vitrification by dimethyl benzene, the samples were

embedded in paraffin. 10μm vertical sections of seeds were cut using Leica manual microtome (Leica FM2235).

A 461bp cDNA fragment of Myb10-D was amplified using specific primers (Myb10-D situ F1: AGCAAGAGGAACCAGCAGTCC; Myb10-D situ R1: TAGCAAAGCCACGCCAACTC) and inserted into the pEASY-Blunt Zero Cloning Vector (TRANS cat.no CB501-01). For the synthesis of antisense and sense RNA probes, the vector was linearization by T7 and T3 RNA polymerase (Roche cat #11031163001; Roche cat #11031163001), respectively. Labeling was following the manufacturer's instruction in the DIG RNA labeling Mixture (Roche, catalog number 1117502910). The sections were observed and imaged with a light stereomicroscope (Leica M165 FC).

S13 Protein expression in *Escherichia coli* and purification

To generate His-tag fusion protein of *Myb10-D*, the full-length cDNA coding sequence of *Myb10-D* was amplified and cloned into the pCold TF DNA vector (TaKaRa) using KpnI and XbaI enzyme sites. The plasmid of the positive clone was then introduced to E.Coli Rosetta (DE3) strain (Weidi CAT# EC1010). Expression of the fusion protein *Myb10-D* -His were induced by 0.5mM IPTG after OD_{600} reached the range of 0.6-0.9, then incubating at 16°C for about 20 hours in the incubator. The fusion protein Myb10-D-His was purified by using Ni-NTA prepacked chromatographic column (Sangon Lot. F618DA0004) according to manufacturer's instruction. To prevent degradation, one mM DTT, one M PMSF or other Protease inhibitors may contain in the disruption buffer and elution buffer. The purification and concentration of purified *Myb10-D* protein were detected by using SDS-PAGE.

S14 Electrophoretic mobility shift assay (EMSA)

Myb10-D cloned into the KpnI and XbaI sites of the pCold TF DNA vector (TaKaRa). The Myb10-D-His fusion protein was purified as described in Ni-NTA prepacked chromatographic column (Sangon Lot. F618DA0004). The NCED promoter sequences was labeled with FAM at the 3'end (Tsingke, Shanghai). Sequences of the primers are listed in Table S1. The probes were mixed with the purified protein at 30°C for 30 min in a reaction containing 0.5 μm FAM-labeled probes 1μl, 1mg/ml salmon sperm DNA

1μl, EMSA Buffer 3.54μl, 1M $MgCl_2$ 0.2μl, 1 M DTT 0.02μl, H_2O 14.24μl. The mixture electrophoresed in 6% polyacrylamide about 80min at constant 100 voltage and luminescence was visualized on the FLA-9000 instrument (FujiFilm, Japan).

S15 Transient expression assay

About 2kb NCED promoter *w*as amplified and fused in frame with the pGreenII-0800-LUC vector digested with KpnI and HindIII to generate the reporter construct *$NCED_{pro}$:LUC*. The CDS of *Myb10-D* was cloned into the PRI101-AN vector to generate the efector constructs *$35S_{pro}$: Myb10-D*. Sequences of the primers are listed in Table S1. Then, the *$35S_{pro}$: Myb10-D* co-transformed with the reporter vector *$NCED_{pro}$:LUC* into d into Agrobacterium GV3101 following the manufacturer's manual (WEIDI, China). The empty vector PRI101-AN and the *$NCED_{pro}$:LUC* were co-transformed as the control. Bacterial suspensions were infiltrated into young, fully expanded leaves of 3-week-old *N. benthamiana* plants using a needleless syringe. Firefly luciferase (LUC) and Renilla luciferase (REN) activity were measured using the Dual Luciferase Assay Reagents (Promega) and a GLOMAX 20/20 luminometer at 2 days after transformation.

S16 One-hybrid assays in yeast

The full-length cDNA sequences of transcription factor *Myb10-D w*as amplified and fused in frame with the GAL4 activation domain in the pGADT7-Rec2 vector digested with Nde I and SacI. The NCED promoter fragments were amplified and inserted into the promoter–pAbAi vector digested with KpnI and XhoI. Sequences of the primers are listed in Table S1. Then, the pGADT7-*Myb10-D* co-transformed with the reporter vector pAbAi-NCED into Y1H Gold yeast following the manufacturer's manual (Clontech, Japan). The empty vector pGADT7-Rec2 and the pAbAi-NCED were co-transformed as the control for mating experiments. DNA-protein interactions were grown on synthetic defined (SD)/-Leu medium with 50mM aureobasidin A (AbA).

S17 Statistical analyses

All the statistical tests used are described in the relevant sections of the manuscript.

Statistical analyses were performed using Microsoft Excel (version 2016). Each experiment represents three independent biological replicates, and all data are expressed as mean ± standard deviation (SD). Differences in metabolism profiling among groups were tested using one-way ANOVA. Z test of multiple percentages with multiple comparisons in Excel was used for statistical analysis of germination ratios among three haplotypes of wheat germplasms. A significant difference among groups or genotypes was claimed at u>1.96 (95% confidence interval) and u>2.58 (99% confidence interval). P values and U values are provided as exact figures where possible.

S18 Code used in this study

Resequencing:

Step1. We need split reference genome sequence. Because CS genome is too large, the genome sequence needs to be segmented. Rice genome can skip this step. We use python program “find_N_section.py” and “get_split_genome.py” (see attached) to split reference genome sequence.

```
python find_N_section.py genome_sequence.fa > genome_N_section.bed
python get_split_genome.py -b genome_N_section.bed -g split_genome.fa
```

Step2. We use bwa (version 0.7.17) to mapping to the reference,and use -K 10000000 to achieve deterministic alignment results.

```
$bwa mem -M -R "@RG\tID:$sample_name\tSM:$sample_name\tPL:Illumina" -t $num_threads -K 10000000 $refer $fq_1 $fq_2 | samtools view -O bam --threads $num_threads $sample_name.bam
```

Step3. We use samtools (version 1.6) to sort, merge and mark duplicate the result of mapping.

```
# Sort bam
$samtools sort --threads $num_threads -O bam -o $sample_name.sort.bam -T ./tmp/ $sample_name.bam
# Merge bam
$samtools merge --threads $num_threads $sample_name.merge.bam $sample_name.sort_1.bam $sample_name.sort_2.bam ...
# Mark Duplicates
```

$samtools rmdup --threads $num_threads $sample_name.merge.bam $sample_name.rmdup.bam

Step4. GATK (version 4.0) was applied to local realignment around indels and recalibrate base quality score.

```
# Local realignment around indels
java -Xmx200g -jar $gatk -T RealignerTargetCreator -R $refer -I $sample_name.rmdup.bam -o $sample_name.intervals
java -Xmx200g -jar $gatk -T IndelRealigner -R $refer -I $sample_name.rmdup.bam -targetIntervals $sample_name.intervals  -o $sample_name.realigned.bam
# Base quality score recalibration
java -Xmx200g -jar $gatk -T BaseRecalibrator -R $refer -I $sample_name.realigned.bam -o $sample_name.table
```

Step5. Use bcftools (version 4.0) to call SNP and get new genome sequence.

```
# call SNP
bcftools mpileup -Ou -f split_genome.fa $sample_name.realigned.bam | bcftools call -vmO z -o $sample_name.vcf.gz
# build vcf file index
bcftools index -c $sample_name.vcf.gz
#get sample genome sequence
bcftools consensus -f split_genome.fa $sample_name.vcf.gz –o; output_$sample_name.fa
```

Step6. Merge the segmented genomic sequences with python program “combine_S_L_v2.py” (see attached).

```
#combine splited sequence
python combine_S_L_v2.py output_$sample_name.fa split_genome.bed $sample_name_genome_sequence.fa
```

“split_genome.bed” format (For example, the sequence of chromosome 1 is divided into seq1S and seq1L, the sequence of chromosome 2 is divided into seq2S and seq2L, and so on.):

```
chr1	0	split_seq1S_length	split_seq1S
chr1	split_seq1S_length	split_seq1S_length+split_seq1L_length	split_seq1L
```

chr2 0 split_seq2S_length split_seq2S

chr2 split_seq2S_length split_seq2S_length+split_seq2L_length split_seq2L

Blast:

We use Blast (version 2.2.28) to get the position of genes in the genomic sequence and gene-to-gene alignment. Set the e-value of the output result to 0.00000001. All the results we selected satisfy the ratio of the number of bases to the length of the gene not less than 90%, and the identity value of result file is greater than 90.

```
blastn -task blastn -num_threads $num_threads -evalue 0.00000001 -db $blast_index -query $yourquery -max_target_seqs $number_chromosomes -outfmt 6 -out $result.psl
```

RNA-seq quant:

We use kallisto(version 0.43.0) to calculate the expression of genes in our material.

```
# build genes index
kallisto index -i $sequences.kallisto_idx $sequences.fa
# quant
kallisto quant --threads 20 -i $sequences.kallisto_idx --output-dir=$your_dir/ $material_1.fq.gz $material_2.fq.gz
```

References

Murray MG, Thompson WF. 1980. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research* **8**: 4321-4325.

~~**Xiao SH. 2002.** Studies on pre-harvest sprouting in wheat. "Resistance of pre-harvest sprouting in wheat" pp123. China Agricultural Science and Technology Press.~~

Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *Peer J* **6**: e4958

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M *et al*. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data.

Nature Genetics **43**: 491-498.

Long XY, Wang JR, Ouellet T, Rocheleau L, Wei YM, Pu ZE, Jiang QT, Lan XJ, Zheng YL. 2010. Genome-wide identification and evaluation of novel internal control genes for Q-PCR based transcript normalization in wheat. *Plant Molecular Biology* **74**: 307-311.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**: e45.

Ishida Y, Hiei Y, Komari T. 2015. High Efficiency Wheat Transformation Mediated by Agrobacterium tumefaciens. In: Ogihara Y., Takumi S., Handa H. (eds) *Advances in Wheat Genetics: From Genome to Field*. Springer, Tokyo, 167-173.

Mieog JC, Howitt CA, Ral JP. 2013. Fast-tracking development of homozygous transgenic cereal lines using a simple and highly flexible real-time PCR assay. *BMC Plant Biology* **13**: 71.

Floková K, Tarkowská D, Miersch O, Strnad M, Wasternack C, Novak O. 2014. UHPLC-MS/MS based target profiling of stressinduced phytohormones. *Phytochemistry* **105**: 147-157.

Schultz AW, Wang J, Zhu ZJ, Johnson CH, Patti GJ, Siuzdak G. 2013. Liquid Chromatography Quadrupole Time-of-Flight Characterization of Metabolites Guided by the METLIN Database. *Nature Protocal* **8**: 451-460.

Liu, L., Zhang, Y., Tang, S., Zhao, Q., Zhang, Z., Zhang, H., Dong, L., Guo, H. and Xie, Q. 2010. An efficient system to detect protein ubiquitination by agroinfiltration in *Nicotiana benthamiana*. The Plant Journal, 61: 893-903.

Janes KA, Wang CC, Holmberg KJ, Cabral K, Brugge JS. 2010. Identifying single-cell molecular programs by stochastic profiling. *Nature Methods* **7**: 311–317.

Supplementary Tables

Table S1 Primers used for KASP, STS, PCR clone, Q-PCR, transgenetic, and transactivation assay.

Table S2 Genotyping and phenotyping of 262 common wheat, 16 Ae. tauschii, and 3 tetraploid wheat lines.

Table S3 Summary of genomic reference sequences data and NGS data obtained in this study (NCBI, PRJNA605937).

Table S4 Transposons in the 2.4 Mb PAV region from CS and AK58.

Table S5 Genes in the 2.4 Mb PAV region from 6 accessions CS, AK58, AL8/78, Robigus, Cadenza, and SHW-L1.

Table S6 The expression of 20 genes in 2.4 Mb PAV region from CM32 and SHW-L1 in 10-30 DPA seeds, and the expression of 3 selected genes from CS and BCScv1 in 29 tissues or stages.

Table S7 Germination percentages of WT and OE lines.

Table S8 Different expression genes that had similar expression patterns with Myb10-D form WT and OE lines in 5-30 DPA seeds.

Table S9 Accumulation of flavonoids and ABA in WT and OE lines.

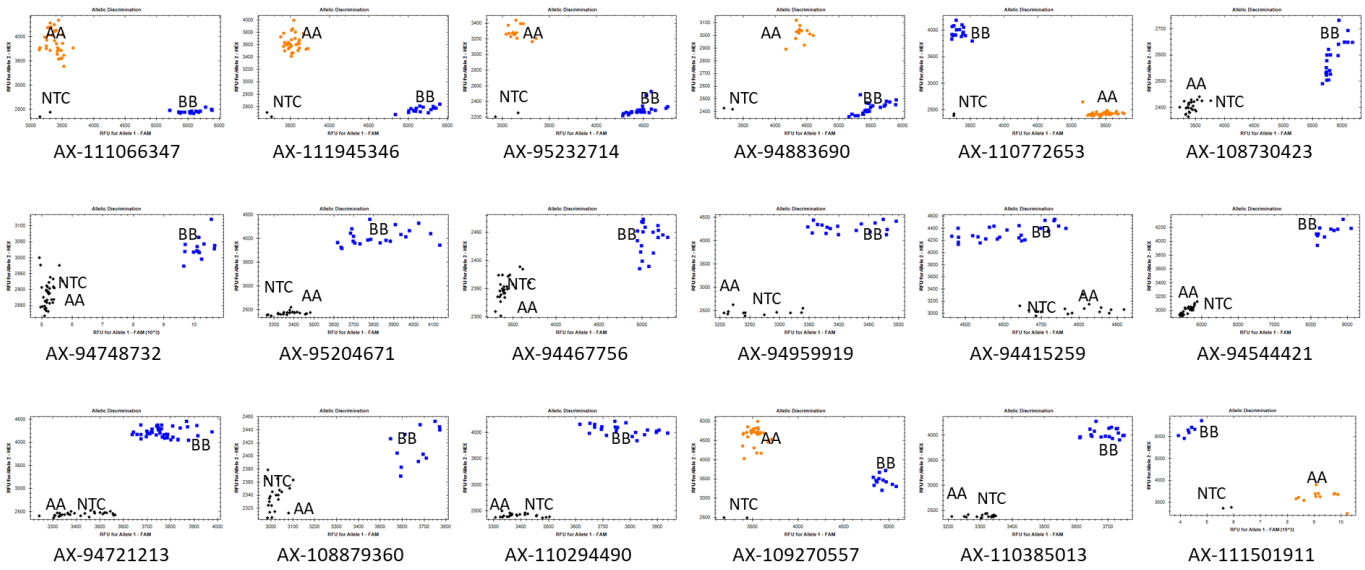

Fig. S1. KASP markers were developed by SNPs in the *QPHS.sicau-3D* (*PHS-3D*) region. The scatter plot with the X(FAM) and Y (HEX) axes show the allelic discrimination of CM32(AA genotypes) or SHW-L1(BB genotypes). The blue, black, and orange dots represent BB genotypes, no signal or AA genotypes, and AA genotypes, respectively.

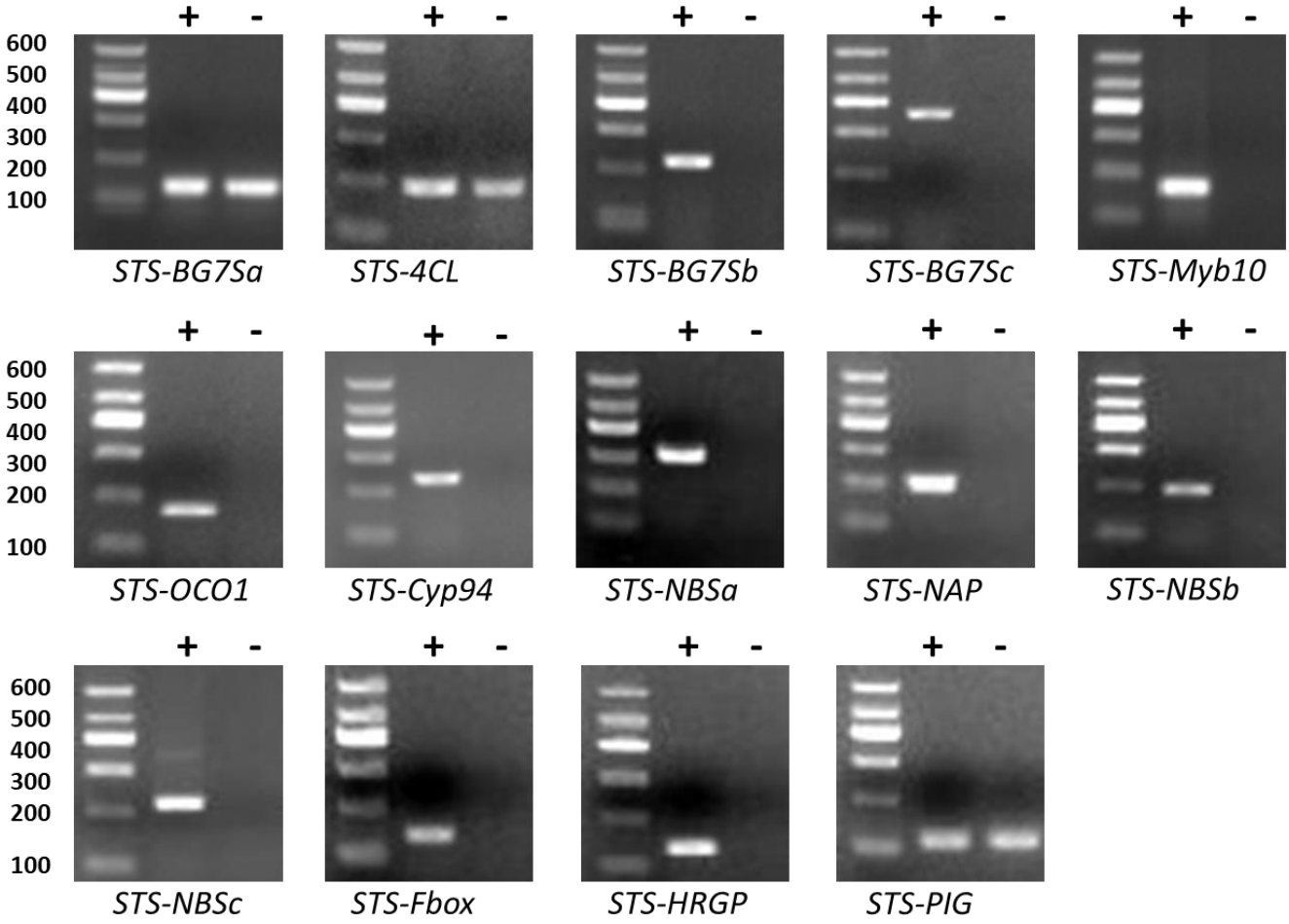

Fig. S2. STS markers were developed in the 2.4 Mb PAG region on 3DL for genotyping wheat lines. These STS markers were based on the sequences of 14 genes/pseudogenes in the PAV region and three genes outside of it. All the PCR fragments were sequenced for validation.

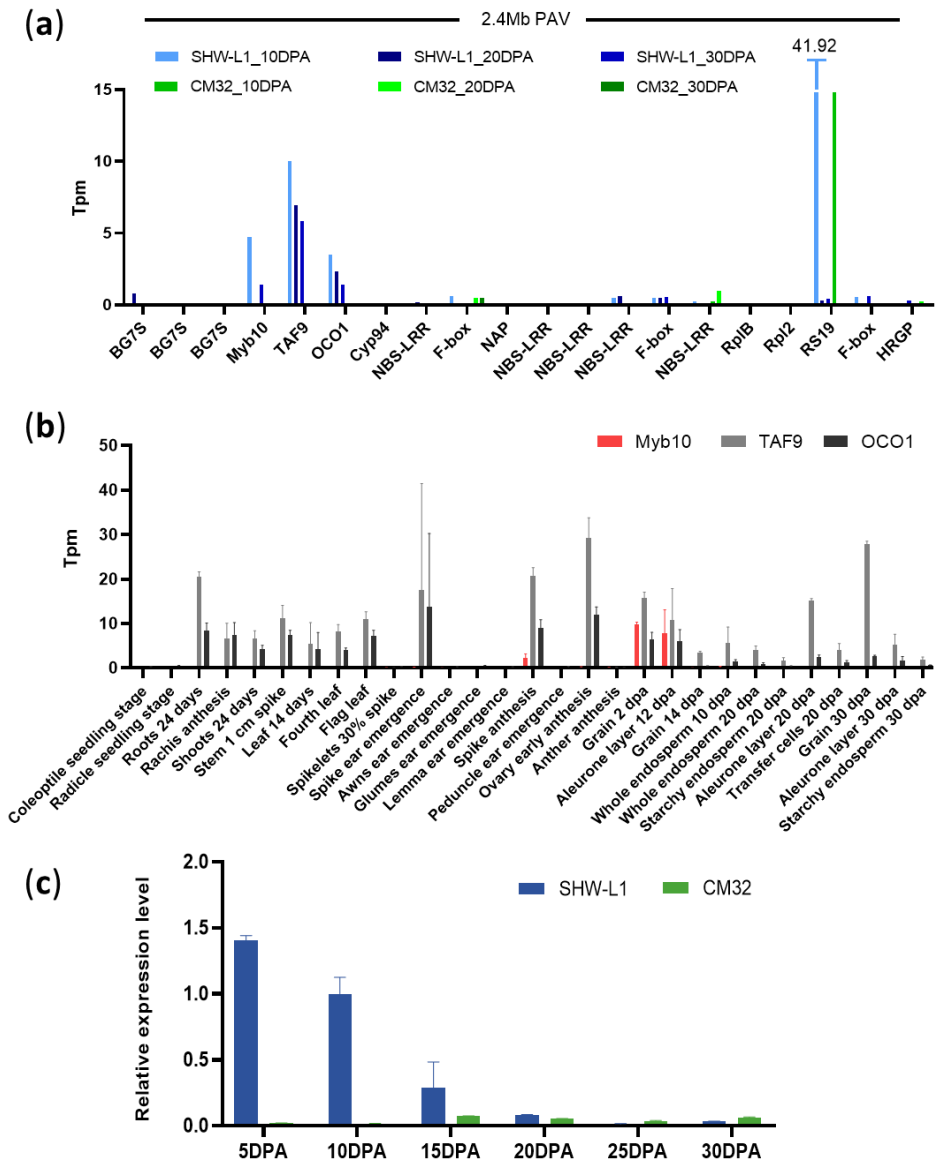

Fig. S3. Expression profiles of genes in the PAV region containing *PHS-3D*. **(a)** Expression of 20 genes in the 2.4 Mb PAV region 10-30 DPA in SHW-L1 and CM32 (RNA-seq). *RS19* (Chloroplastic 30S ribosomal protein S19) is highly expressed in both SHW-L1 and CM32 that may result from a mismatch of reads from other chloroplastic orthologs. **(b)** Expression of *Myb10*, *TAF9*, and *OCO1* across 29 tissues and stages in CS and BCScv1 (RNA-seq). **(c)** Expression of Myb10 5-30 DPA in SHW-L1 and CM32 (qPCR). Standard deviation errors bars was given in the graphs.

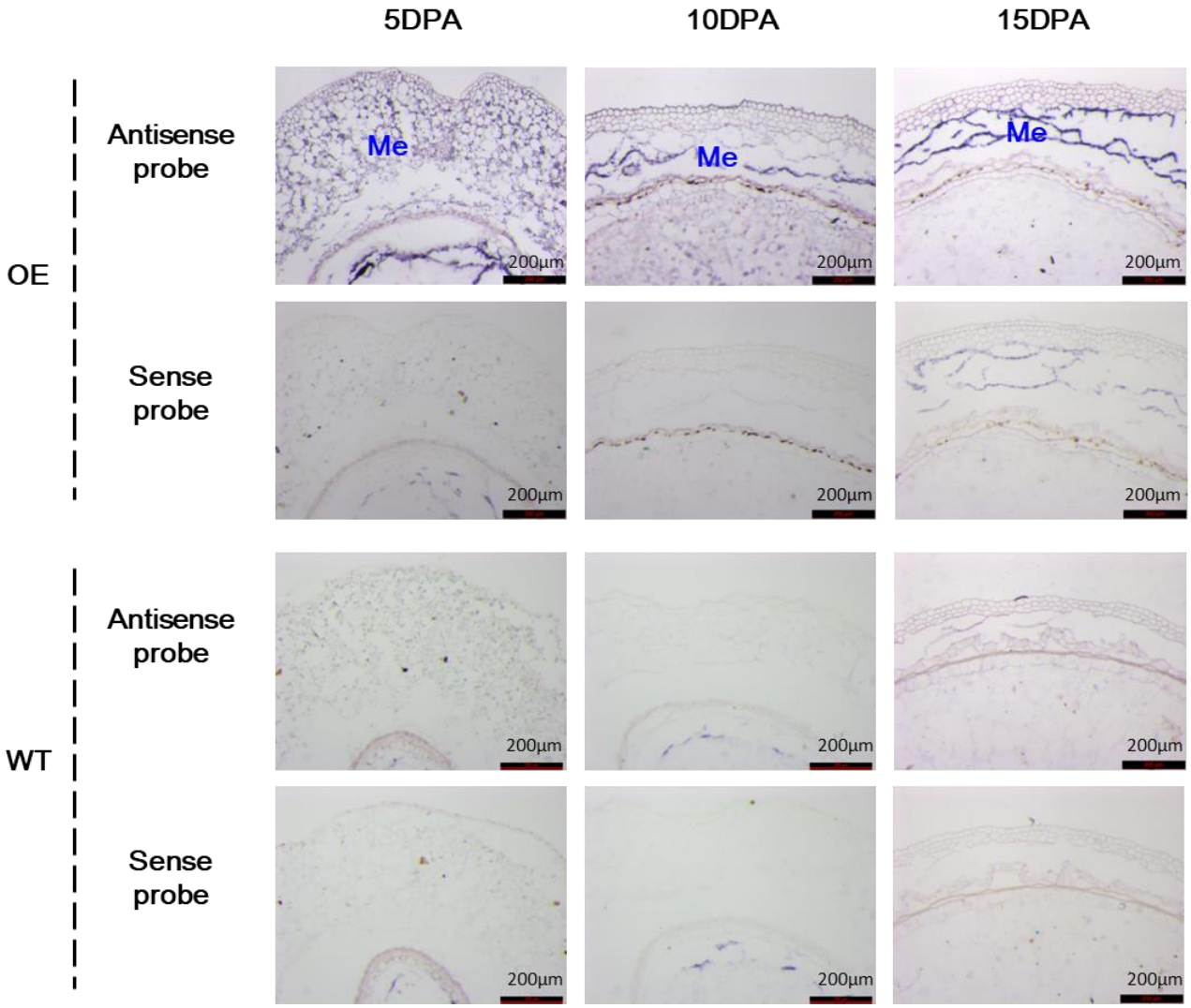

Fig. S4. RNA in situ hybridization of *Myb10-D* in transversal sections 5-15 DPA indicates its expression in the mesocarp of the inner pericarp. The positive signals can mainly be observed in the mesocarp (Me). Scale bars are shown in each figure. Bars=200μm

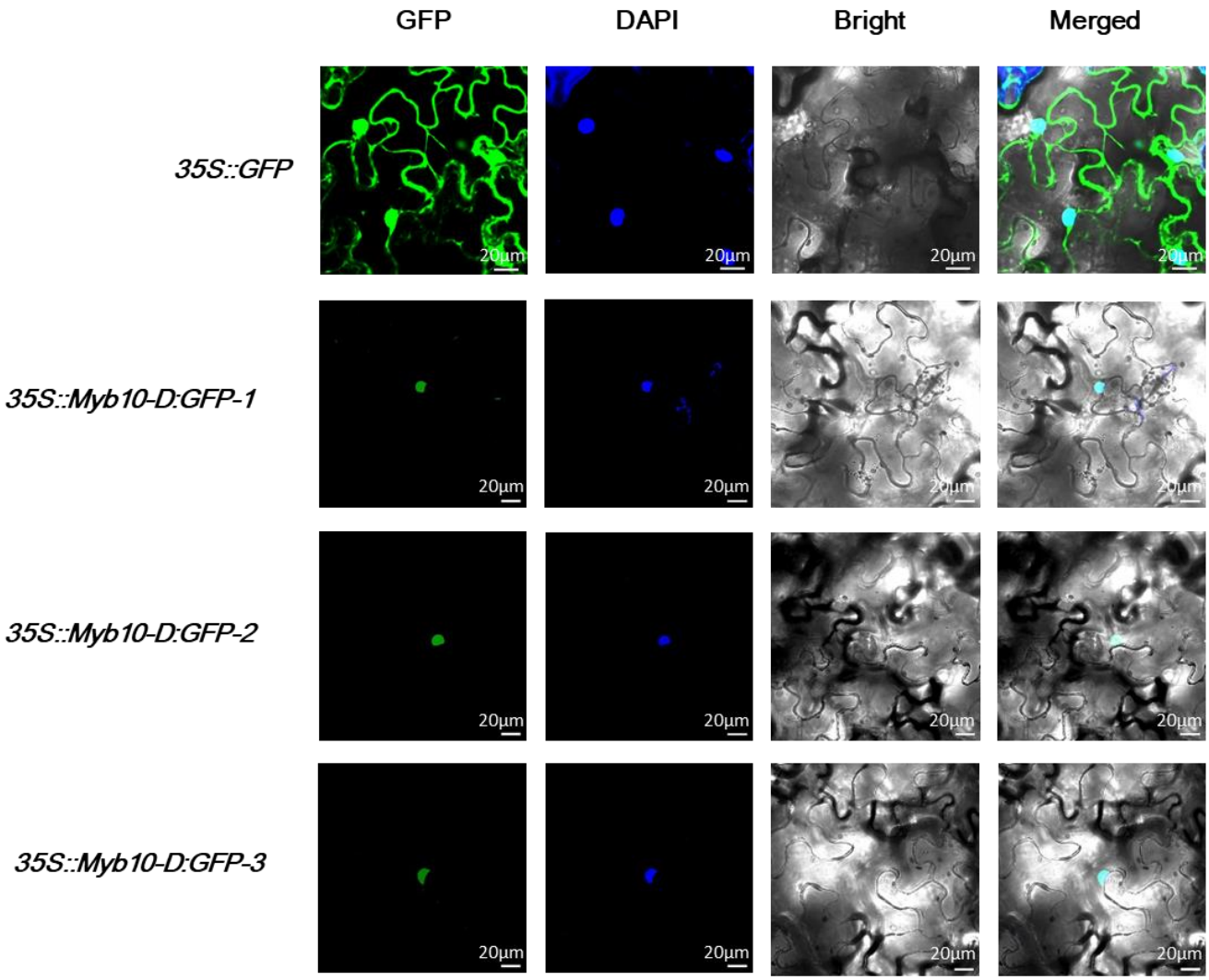

Fig. S5 Subcellular localization of Myb10-D in the nucleus. 35S::Myb10-D:GFP fused proteins were transiently expressed in N. benthamiana. Subcellular localization analysis using confocal microscopy in three replicates. DAPI=nuclear localization marker. Bars=20μm for each panel.

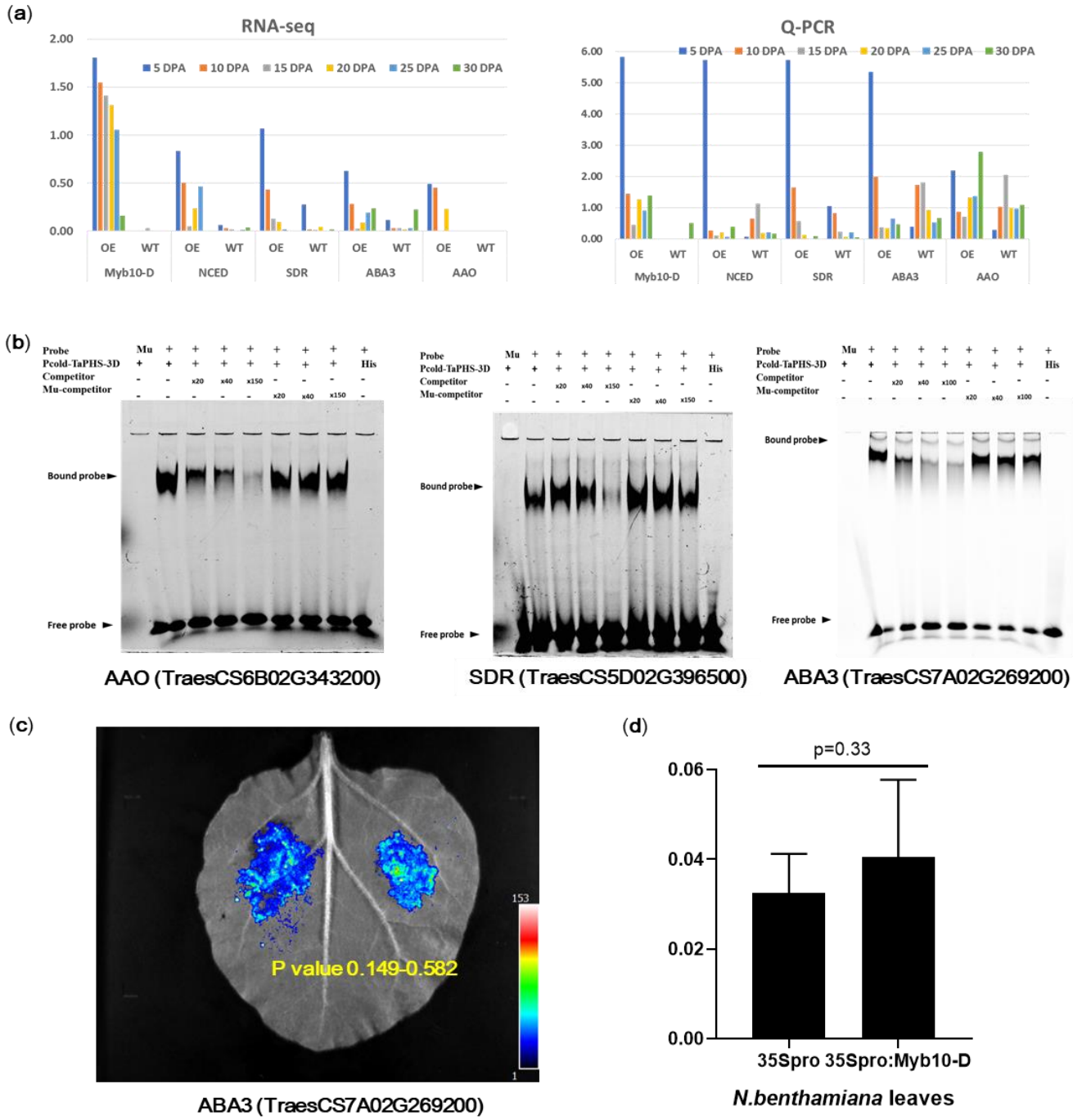

Fig. S6 Characterization of binding sites based on the 2-kb promoter regions of potential *Myb10-D* regulated genes in the MEP pathway. **(a)** Expression profiles of *Myb10-D*, *NECD*, *SDR*, *ABA3*, and *AAO* 5-30 DPA. **(b)** EMSA of *Myb10-D* with probes containing the binding motifs in the promoters of *SDR*, *ABA3*, and *AAO*. **(c)** Dual-luciferase assay of *Myb10-D* in the promoters of *ABA3* in *N. benthamiana* leaves. **(d)** None significant differences were detected in *N. benthamiana* leaves for *ABA3*. Standard deviation errors bars was given in the graphs.