A Single Nucleotide Polymorphism in an R2R3 MYB Transcription Factor Triggers *ms6 (Ames1)* Male Sterility in Soybean

Junping Yu
Northwest University
Guolong Zhao
Jilin Academy of Agricultural Sciences
Wei Li
Northwest University
Ying Zhang
Jilin Academy of Agricultural Sciences
Peng Wang
Northwest University
Aigen Fu
Northwest University
Limei Zhao
Jilin Academy of Agricultural Sciences
Chunbao Zhang
Jilin Academy of Agricultural Sciences
Min Xu (xumin@nwu.edu.cn)
Northwest University https://orcid.org/0000-0002-7206-0059

Original Article

Keywords: Glycine max, Anther, Male sterile, GmMS6, MYB transcription factor

Posted Date: February 12th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-188606/v1

Version of Record: A version of this preprint was published at Theoretical and Applied Genetics on July 28th, 2021. See the published version at https://doi.org/10.1007/s00122-021-03920-0.

Abstract

Soybean [*Glycine max* (L.) Merr.] is an important crop providing vegetable oils and proteins. Increasing demand on soy products heightens the urgency of soybean yield improvement. Hybrid breeding with male sterility system is an effective method to improve crop production. Cloning of genic male sterile (GMS) gene combined with biotechnology method can contribute to constructing GMS-based hybrid Seed Production Technology (SPT) to promote soybean performance and yield. In this research, we identified a soybean GMS locus, *GmMS6,* by combining bulked segregant analysis (BSA)-sequencing and map-based cloning technology. *GmMS6* encodes an R2R3 MYB transcription factor, whose mutant allele in *ms6* (*Ames1*) harbors a single nucleotide polymorphism (SNP) substitution, leading to the 76th Leucine to Histidine change in the DNA binding domain. Phylogenetic analysis demonstrates GmMS6 is a homolog of Tapetal Development and Function 1 (TDF1)/MYB35 that is an anther development key factor co-evolved with angiosperm. It has a recently duplicated homolog GmMS6LIKE (GmMS6L), both of which can rescue the male fertility of *Arabidopsis* homologous mutant *attdf1* while GmMS6^{L76H} cannot, denoting that both proteins are functional and L76 is a critical residue for TDF1's function. However, compared to anther specific expressed *GmMS6, GmMS6L* is constitutively expressed at a very low level, explaining deficiency of GmMS6 alone causes pollen abortion. Moreover, the expression levels of major regulatory and structural genes for anther development are significantly decreased in *ms6*, unveiling that GmMS6 is a core transcription factor regulating soybean anther development.

Introduction

Soybean (*Glycine max*) is the major crop to provide plant proteins and oils in food supply, but has a relatively low yield compared to other major crops. Hybrid breeding technology that significantly improves crop yield has a great potential in soybean seed production (Kim and Zhang 2018; Palmer et al. 2001). HybSoy1, the first officially approved and commercially applicable hybrid soybean, could increase the yield by 20.8% (Zhao et al. 2004). In hybrid breeding system, male sterile lines are indispensable for avoiding the time-consuming and tedious artificial emasculation process.

The general hybrid breeding systems used today are three-line and two-line systems derived from cytoplasmic male sterility (CMS) and environmental sensitive genic male sterility (EGMS), respectively (Kim and Zhang 2018). However, both systems have some limitations to hinder their broad applications. For example, it is difficult for CMS to find suitable restorer lines. Some CMS cytoplasm types even have negative effects on crop performance such as leading to disease susceptibility (Levings 1990). As to EGMS, the sterility is relatively unstable for its high reliance on the environmental conditions, which severely affects hybrid purity sometimes (Chen et al. 2011). Although stable genic male sterility (GMS) can overcome these defects, lack of maintainer line has restricted its usage in hybrid production for a long time until the seed production technology (SPT) is developed recently (Perez-Prat and van Lookeren Campagne 2002; Weber et al. 2009). The main idea of this technology is to create a transgenic-based maintainer line in a recessive sporophytic male sterile mutant (*ms*) background by introducing a gene cluster besides the resistance gene for screening transgenic lines, and the transgenic component in SPT

maintainer line is kept in heterozygote status. The gene cluster is composed by at least three fundamental genes (Wu et al. 2016). The first one is the wild-type *MS* allele controlled by its native promoter for rescuing *ms*'s detrimental effects on anther sporophytic cells. The 2nd one is a male gametophyte-killer gene, for killing the microspores carrying the transgenic component, so that only non-transgenic *ms* pollens are viable for hybrid production (Chang et al. 2016; Song et al. 2020). The 3rd one is a phenotypic reporter gene for monitoring the purity of obtained *ms* seeds, such as fluorescence gene expressed in aleurone layer for monocots rice and maize (Chang et al. 2016; Zhang et al. 2018b) or anthocyanin synthesis gene expressed in early seedling stage for dicots tomato (Du et al. 2020).

SPT system broadens the germplasm choices of parental lines to breed hybrids of superior heterosis, reduces the risk caused by weather changes, and is regarded as the third generation of hybrid technology. However, it has not been applied in soybean as lack of cloned GMS gene. So far, 13 non-allelic genic loci distributed on 7 different chromosomes have been reported to condition anther development in soybean, including *ms1-ms9*, *msp, msMOS*, *mst-M*, and ms_{NJ} (Yang et al. 2014; Zhao et al. 2019; Nie et al. 2019; Thu et al. 2019). Mutations at these loci all confer recessive sporophytic male-sterile phenotype. Among these mutants, *ms6* displays a stable non-pollen phenotype, making it an ideal material for developing soybean SPT system. There are two independent and spontaneous *ms6* mutants maintained as heterozygotes in Soybean Genetic Type Collection as T295H (*ms6* (*Ames1*)*/+*) (Skorupska and Palmer 1989) and T354H (*ms6* (*Ames2*)*/+*) (Ilarslan et al. 1999), respectively. Comparative microscopic study of the anther development of fertile and sterile plants from T354H has showed that the cytological abnormalities of *ms6* anther firstly appear at microspore mother cell (MMC) stage on tapetal and parietal layers, which possess more vacuoles in cells compared to fertile anther (Ilarslan et al. 1999). Then, tapetum in *ms6* anther is severely degenerated, forming condensed tissues from meiosis to late tetrad stage, and completely degraded in the late microspore stage when tapetum in fertile anther just starts enlargement and vacuolation (Ilarslan et al. 1999). By contrast, the parietal layer in *ms6* anther keeps enlarging during the later development stages and shows completely vacuolated at the end, while it remains in a rather consistent shape in fertile anther (Ilarslan et al. 1999). The reproductive cells in *ms6* anther show aberrations since telophase II. Meiocytes fail cytokinesis and form partially separated microspores, which would be completely collapsed in the late microspore stage when the fertile microspores are processing the first mitosis (Ilarslan et al. 1999). Similar phenomenon was observed during the microsporogenesis in *ms6* mutant from T295H (*ms6* (*Ames1*)*/+*) (Skorupska and Palmer 1989). Multi-nucleic microspores are generated after meiosis, and they are completely crushed later so that no pollen is produced in the sterile plants (Skorupska and Palmer 1989).

The *ms6* locus has been mapped into a 3.7 Mb region on chromosome 13 (Chr13) between two SSR markers Satt030 and Satt149 (Yang et al. 2014), closely linked to a flower-color gene *W1* (Skorupska and Palmer 1989; Lewers and Palmer 1993; Ilarslan et al. 1999). In this study, we further narrowed down the genetic region of *ms6* and identified the mutation corresponding to *ms6* via BSA-sequencing, map based cloning and complementation experiments. It is a missense mutation in the gene, named as *GmMS6*, which encodes a homolog of TDF1, an R2R3 MYB transcription factor critical for anther development in

Arabidopsis and rice by regulating tapetal layer degeneration (Zhu et al. 2008; Cai et al. 2015). We revealed that although TDF1 has two functional paralogs in soybean, GmMS6 (GmTDF1a) and GmMS6-like (GmTDF1b), GmTDF1a is the major one regulating soybean anther development. The expression of anther development factors related to TDF1 regulatory pathway were also compared in WT and *ms6*, revealing that this genetic pathway is conserved but more complicated due to recent whole genome duplication in soybean. The results from our study not only provide new insights into the regulatory network in soybean anther development, but also turn *ms6* mutant to be a practicable material for SPT system to facilitating hybrid seed production in soybean.

Materials And Methods

Primers

Primers used in present study were listed in Supplementary Table 1.

Plant materials and growth conditions

The *ms6* mutant used in this study is derived from T295H (PI 533601, *ms6* (*Ames1*)/+), which was achieved from the collection of National Plant Germplasm System (NPGS) in United States. Allele *ms6* (*Ames1*) is referred as *ms6* hereafter. The BC_5F_2 segregating population was developed for narrowing down the genetic region of *ms6*, by using T295H as *ms6* donor and a wild-type (WT) cultivar 'JiuB', from Jilin, China, as a recurrent male parent. The mapping population was planted in the farm of Fanjiatun, Jilin in summer. For cytological and morphological studies, soybean materials were grown in pots (two plants per pot) outdoors in summer and in the greenhouse in winter at 28°C with the photoperiod of 16 h light /8 h dark, in Xi'an, China.

Arabidopsis and *Nicotiana benthamiana* plants were grown in soil in the greenhouse at 22°C with the photoperiod of 16 h light /8 h dark, in Xi'an, China. The *Arabidopsis* germplasms used in this study were WT Columbia (Col), heterozygous *attdf1* (in Col background), transgenic lines in homozygous *attdf1* background carrying desired transgenes for the complementary experiment. As noted, *attdf1* mutant used here was obtained by the introgression of *attdf1* locus from Landsberg *erecta*-0 into Col background through several rounds of backcrossing.

Morphological and cytological analysis

For general morphological observation of *ms6* anthers, soybean flowers one day before blooming were collected from fertile and sterile descendants of T295H. Stamens were dissected and imaged under the stereo microscope Nikon SMZ25. Pollens were squeezed out, stained with 1% I_2-KI solution, and photographed under light microscope Leica DM2500. For *Arabidopsis*, mature anthers before anthesis collected from WT, *attdf1* mutant, and various transgenic lines were stained with Alexander staining buffer (Peterson et al. 2010), and photographed under Leica DM2500.

For cytological analysis, flowers at late tetrad and late microspore stages were collected from fertile and sterile descendants of T295H, and immediately immersed into the FAA fixation solution. After dehydration, flower samples were imbedded into resin with Technovit H7100-GMA kit (Heraeus Kulzer, Germany), following manufactory instruction, and sliced into 2-μm transverse sections with Leica RM2265. Sections on slides were stained with 0.5% toluidine blue staining buffer, and imaged under Leica DM2500 after sealed.

DNA extraction, BSA-sequencing (BSA-seq), and fine mapping

Genomic DNA samples were extracted from the young leaves with the Nuclean Plant Genomic DNA Kit (CWBIO, China) for regular PCR analysis and BSA-seq experiment. For BSA-seq analysis, two bulks were constructed from the BC_5F_2 mapping population. One was composed of 20 homozygous WT plants and the other 20 homozygous *ms6* plants. Genomic DNA isolated from each bulk was fractioned to build a 350-bp pair-end sequencing library, and sequenced on Illumina Hiseq PE150 platform at Novogene Company (China). SNP (single nucleotide polymorphisms) and InDel (insertions-deletions) of each bulk were annotated using Wm82.a2.v1 genome as the reference. The SNP-index of each bulk was calculated as previously described (Takagi et al. 2013). The SNP-index differences between two bulks, Δ(SNP-index), were calculated and plotted against their genomic positions.

Polymorphic SSR markers in the genetic window of *ms6* locus identified by BSA-seq were further used to screen the individual *ms6* plants in the BC_5F_2 population via canonical PCR. The fragments amplified with SSR primers were resolved on 8% polyacrylamide gel in 1xTAE buffer by electrophoresis, and visualized by silver staining method (Bassam et al. 1991). The genetic map was constructed from the data with MAPMAKER 3.0 (Lander et al. 1987).

Bioinformatics and phylogenetic analysis

The conserved structural domain in GmMS6 was predicted by SMART (http://smart.embl-heidelberg.de/), showing it contained a typical R2R3 MYB DNA-binding domain. The conservancy in R2 motif was further analyzed by aligning the DNA-binding domain of GmMS6/GmTDF1a to the ones of well-characterized MYB proteins in *Arabidopsis*, and the results were drawn with Bioedit software (Tom Hall). By using BLASTP, the homologs of GmMS6 in *Glycine max* (GmMS6L) and *Arabidopsis* (AtTDF1) were identified in NCBI Database. The sequence conservancy of these three proteins was assessed by multiple sequence alignment with Bioedit software (Tom Hall).

For phylogenetic analysis, we used the protein sequences of GmMS6, GmMS6L, and homologs of TDF1 in 15 representative species from different land plant evolutionary lineages, including five dicots (*G. max*, *Medicago truncatula*, *Vitis vinifera*, *Arabidopsis thaliana* and *Solanum lycopersicum*), six monocots (*Oryza sativa*, *Zea mays*, *Sorghum bicolor*, *Ananas comosus*, *Musa acuminate* and *Zostera marina*), one basal angiosperm (*Amborella trichopoda*), two gymnosperm species (*Ginkgo biloba* and *Pinus taeda*), one lycophyte (*Selaginella moellendorffii*), and one moss (*Physcomitrella patens*). The sequences of

TDF1 were retrieved from NCBI with BLASTP by using AtTDF1 (NP 189488.1) and OsTDF1 (XP 015630216.1) as query peptides, and the one with highest bit-score in each species was selected. AtMYB80 (NP 200422.1) and OsMYB80 (XP 015635420.1) were used as outgroup sequences. All the protein sequences were subject to multiple sequence alignment analysis with the ClustalW2 algorithm, and the phylogenetic tree was constructed by neighboring-joint (NJ) method with bootstrap resampling (1000 replicates) by using MEGA 6 (Tamura et al. 2013).

RNA extraction, RT-PCR, and qRT-PCR analysis

Total RNA was extracted from desired tissues with RNAprep Pure Plant Kit (Tiangen, China). After removing genomic DNA contamination with TURBO DNA-free™ Kit (Invitrogen, United States), 1 μg of RNA was reversely transcribed into the cDNA by PrimeScript™ II 1st Strand cDNA Synthesis Kit (Takara, Japan). Sequential PCR was conducted with rTaq DNA polymerase (Takara, Japan) for analyzing expression level and with 2x PrimeSTAR Max Premix (Takara, Japan) for cloning purpose following manufactory instructions. For quantitative RT-PCR, cDNA samples after 1:10 dilution were used as templates. The PCR reactions were conducted as previously described (Zhang et al. 2018a). For assessing the expression patterns of *GmMS6* and *GmMS6L*, relative gene expression levels were calculated by using the $2^{-\Delta Ct}$ method. For assessing the differential expression of specific genes in WT and *ms6* young flowers, fold changes in gene expression were calculated by using the $2^{-\Delta\Delta Ct}$ method. All data were normalized against the expression level of *GmActin11* (*Glyma.18g290800*). For each sample, three replicates were performed.

Subcellular localization analysis

The full length coding sequence (CDS) without stop codon of *GmMS6*, *GmMS6L*, and mutant *GmMS6*^{L76H} were in frame cloned into the *Xba*I site upstream of the *GFP* gene in the binary vector of pLM-*35S-GFP* to create pLM-*35S-GmMS6-GFP*, pLM-*35S-GmMS6L-GFP*, and pLM-*35S-GmMS6*^{L76H}*-GFP*. Vectors were transformed into *Agrobacterium tumefaciens* strain *GV3101* and infiltrated into 4-week-old *N. benthamiana* leaves. GFP signals were observed and imaged at 48 hours post infiltration under the Olympus Fluoview FV1000 confocal laser scanning microscope (Olympus, Japan).

Transactivation activity assay in Yeast

Full-length CDSs corresponding to GmMS6, GmMS6^{L76H} and GmMS6^{DBD} (1-191 aa of GmMS6, DNA binding domain) were in frame cloned into the pGBKT7 vectors downstream of the CDS of GAL4-BD, respectively. The obtained vectors of pGBKT7-GmMS6, pGBKT7-GmMS6^{L76H}, and pGBKT7-GmMS6^{DBD} were subsequently transformed into *Saccharomyces cerevisiae* stain *AH109* via one-step transformation method (Chen et al. 1992). After selected on synthetic dropout medium lack of trypsin (SD/-Trp), the positive colonies were diluted into the same concentration with autoclaved ddH_2O. Then, 5 μL of cell suspensions were placed on selective medium SD/-Trp/-His, and grown for 3–4 days under 30°C to evaluate the activation activities of target proteins.

Complementary analysis

The 817-bp *AtTDF1* promoter, reported previously (Zhu et al. 2008), were cloned into the *pCAMBIA1301* via *Kpn*I and *Xba*I to get pCAMBIA1301-*AtTDF1pro* vector. The CDSs of *GmMS6*, *GmMS6*^{L76H}, *AtTDF1*, *AtTDF1*^{L46H}, and *GmMS6L* were cloned into pCAMBIA1301-*AtTDF1pro* downstream of *AtTDF1* promoter through *Xba*I and *BstE*II to acquire pCAMBIA1301-*AtTDF1pro-GmMS6*, pCAMBIA1301-*AtTDF1pro-GmMS6*^{L76H}, pCAMBIA1301-*AtTDF1pro-AtTDF1*, pCAMBIA1301-*AtTDF1pro-AtTDF1*^{L46H} and pCAMBIA1301-*AtTDF1pro-GmMS6L*. *A.tumefaciens* strain *GV3101* carrying these vectors were used to transform *Arabidopsis attdf1* heterozygote plants by floral-dip method (Clough and Bent 1998). T1 transgenic seeds were selected by sowing on 1/2 MS medium supplemented with 1% sucrose and 20 mg/L hygromycin. After verified by PCR, T1 transformants were transplanted to soil for further growth. The genotype of native *attdf1* locus in each T1 transgenic plant was evaluated with a CAPS (Cleaved Amplified Polymorphic Sequences) marker based on the mutated site. The transgenic plants in homozygous *attdf1* background were further scored the fertility.

Result

Phenotypic characterization of *ms6* mutant

The offspring of heterozygous *ms6* plants (T295H) were planted in a greenhouse condition. During vegetative stage, all plants grew well just like wild types (WT), while in reproductive stage about a quarter of plants (*ms6*) were male-sterile and unable to develop pods after blooming (Fig. 1a). Compared to the anthers in WT plants, the anthers of *ms6* plants were more whitish and shrinking (Fig. 1b). Pollen grains released from WT anthers were round and turned dark blue after stained by I_2-KI solution (Fig. 1c), but no pollen grains were produced in *ms6* sterile anthers (Fig. 1d). These results are consistent with the previous report (Skorupska and Palmer 1989).

WT and *ms6* flowers at late tetrad and late microspore stages were cross-sectioned and observed under light microscopy. At late tetrad stage, WT anther wall was composed of 5-layers, including epidermis, endothecium, middle layer, parietal layer, and tapetum from outside to inside, in which the cytoplasm of tapetum cells is highly condensed; meanwhile, callose surrounding the tetrads in locule appeared to start degeneration (Fig. 1e). Anthers at same stage in *Arabidopsis* and rice show similar cytological features except that their anther walls lack of the parietal layer (Sanders et al. 1999; Zhang et al. 2011). On the other hand, *ms6* anther had radically enlarged and highly vacuolated parietal and tapetal layers; in the locule, callose encompassed partially or non-separated irregular microspores with multiple nuclei, indicating that cytokinesis II of meiocytes was abnormal (Fig. 1f). This is similar to the phenotype of *ms6* (*Ames2*) (T354H), but tapetal layer of *ms6* (*Ames2*) was degenerated more rapidly and was almost completely degraded at this stage (Ilarslan et al. 1999). By the time of early pollen stage, enlarged pollens with thick walls were observed in the locule of WT anther, and the anther wall was composed by an epidermis, an enlarged endothecium, and a narrow parietal layer with attachment of remnant tissue from degraded tapetum cells (Fig. 1g). Comparatively, in *ms6* anther, tapetum cells were completely dissolved

and pollens were crushed, while parietal layers were abnormally vacuolated and swollen, similar to the situation reported in *ms6* (*Ames2*) (T354H) (Fig. 1h; Ilarslan et al. 1999).

A SNP mutation of *Glyma.13g066600,* an R2R3-MYB transcription factor encoded gene, is likely responsible for the male sterility in *ms6*

The *ms6* loci was mapped previously within a 3.72 Mb region on Chr13 between two SSR markers Satt149 (Chr13:13,134,055 bp) and Satt030 (Chr13:16,855,019 bp) (Yang et al. 2014). To narrow down the region, WT and *ms6* bulks were constructed from a BC_5F_2 mapping population derived from T295H and 'JiuB', and subject to BSA-seq analysis. Plotting the ΔSNP-index values between two bulks against their genomic positions showed that *ms6* was associated with a 1.5 Mb region (15,853,267 – 17,349,424 bp) on Chr13 (Fig. 2a), consistent to the previously reported interval (Yang et al. 2014). In addition, 249 variations between two bulks were identified in this region, including 214 SNP and 35 InDel.

Subsequently, a fine mapping was conducted with 328 individual plants in the BC_5F_2 mapping population with 9 polymorphic SSR markers identified in the 1.5 Mb interval, including BARCSOYSSR-13-0243, BARCSOYSSR-13-0244, BARCSOYSSR-13-0245, BARCSOYSSR-13-0249, BARCSOYSSR-13-0257, BARCSOYSSR-13-0259, BARCSOYSSR-13-0275, BARCSOYSSR-13-0277, and BARCSOYSSR-13-0283 (Fig. 2b; Fig. S1). Finally, *ms6* was restricted to a 255 kb region (Chr13: 16,428,596 – 16,683,664 bp) between SSR markers BARCSOYSSR-13-0259 and BARCSOYSSR-13-0275 (Fig. 2b), which harbored 23 annotated genes. According to the BSA-Seq results, amongst these genes, only *Glyma.13g066600* had an SNP (T-to-A, Chr13: 16,641,429 bp) in the CDS region (Fig. 2c). The SNP in *ms6* destroyed an *Mse*I restriction site (TTAA) that was present in WT sequence (Fig. 2c).

To verify the mutation, a 126-bp region covering the SNP site was amplified by PCR from homozygous WT (+/+), heterozygotes (+/*ms6*), and homozygous *ms6* (*ms6*/*ms6*), and subsequently digested with *Mse*I. As expected, the amplicon of WT was cleaved to a 99-bp band and a 27-bp band that was invisible on the agarose gel. Comparatively, the amplicon from *ms6* plants could not be digested at all, while about half of the amplicons from heterozygotes could be cut (Fig. 2e), evidencing the mutation of T to A at Chr13:16,641,429 bp in *ms6*. Moreover, the 126-bp segment could be used as a CAPS marker to differentiate the plant genotype at *ms6* locus at any stage.

Glyma.13G066600 encodes a typical R2R3-MYB transcription factor with two MYB motifs close to the N terminus served as DNA binding domain (Fig. 2d). This protein showed a strong homology to TDF1, amino acid sequence exhibiting 48% identity and 59% similarity to AtTDF1 in *Arabidopsis* (Fig. S2). TDF1 is known as a key transcription factor in regulating tapetum development and function. Null mutant of *TDF1* in *Arabidopsis* and rice both promoted vacuolization in tapetal cells and suppressed the degradation of the callose surrounding the tetrads, resulting in squeezed microspores and no pollen formation (Zhu et al. 2008; Cai et al. 2015), which was similar to *ms6*. The SNP in *Glyma.13g066600*

CDS in *ms6* leads to the amino acid substitution of leucine to histidine at the residue 76 (L76H) (Fig. 2d), which is a well-conserved residue in the R2 MYB motif (Fig. 2f). These results suggested that mutation at *Glyma.13g066600* was responsible for the male sterility of soybean *ms6*, and therefore its wild-type allele was termed as *GmMS6*.

GmMS6 is a homolog of TDF1 that is only present in angiosperm

Blast search showed that GmMS6 had a homolog GmMS6L with 92% amino acid sequence identity, encoded by *Glyma.19g017900* (Supplementary Fig. S2). Similar to AtTDF1, GmMS6L doesn't have the N30 extension in the sequence (Supplementary Fig. S2). To further confirm the relationship between GmMS6 and TDF1, we conducted a phylogenetic analysis to GmMS6, GmMS6L, and the homologs of AtTDF1 and OsTDF1 in different land plant evolutionary lineages. AtMYB80 and OsMYB80 were used as outgroup sequences for MYB80 is the closest homolog of TDF1, also known as MYB35 (Dubos et al. 2010).

The result showed that TDF1 homologs in angiosperm were clustered into a monophyletic group with two branches. All the dicots were grouped in a branch with the basal angiosperm species *A. trichopoda*, while all the monocots were grouped in the other branch with the basal or near basal monocot species *A. comosus* and *Z. marina* (Fig. 3). Comparatively, the TDF1 homologs with highest bit-score in lycophyte *S. moellendorffii* and moss *P. patens* were clustered with MYB80 and those in gymnosperm species, *G. biloba* and *P. taeda*, had even a further evolutionary relationship. These data suggested that TDF1 is only present in angiosperm and has diverged at a very early stage in angiosperm evolution before monocots and dicots were differentiated.

In the phylogenetic tree, GmMS6 and GmMS6L were positioned together in the TDF1 clade, evidencing that they were homologs of TDF1 and evolved from a recent duplication (Fig. 3). It is interesting to know whether L76H in GmMS6 ($GmMS6^{L76H}$) could lead to male sterility, and if so, why GmMS6L fails to compensate the GmMS6's function in *ms6* mutant.

GmMS6 but not $GmMS6^{L76H}$ could complement the *attdf1* male sterile phenotype in *Arabidopsis*

The function TDF1 was likely conserved in all the angiosperm species. Firstly, phylogenetic results showed that TDF1 in angiosperm were clustered into a monophyletic group (Fig. 3). Secondly, depletion of TDF1 in *Arabidopsis* and rice caused similar detrimental effects on tapetum development as lack of GmMS6 did (Zhu et al. 2008; Cai et al. 2015). Therefore, we speculated that GmMS6 could substitute the role of AtTDF1 in *Arabidopsis* although GmMS6 is 30 amino acids longer at the N terminus, while $GmMS6^{L76H}$ could not.

To verify this, we firstly transformed heterozygous *attdf1* (+/*attdf1*) plants with the CDSs of *AtTDF1*, *GmMS6*, *GmMS6L*, and *GmMS6*^{L76H} driven by native promoter *AtTDF1p*. T1 transgenic plants in homozygous *attdf1* background were subsequently assessed for anther fertility. As it turned out, most *attdf1* transformed with *AtTDF1* (14/20, rescued/transformants), *GmMS6* (19/26, rescued/transformants) and *GmMS6L* (8/10, rescued/transformants) were fully complemented, producing functional pollens and elongated siliques as wild type did (Fig. 6a-e, 6h-l, 6o-p), and demonstrating that both GmMS6 and GmMS6L are functional TDF1. Therefore, we termed these two proteins as GmTDF1a and GmTDF1b, correspondingly. By contrast, *AtTDF1p:GmMS6*^{L76H} (0/25, rescued/ transformants) failed to complement the *attdf1*'s sterility (Fig. 6a-b, 6f, 6h-i, 6m, 6r), confirming that GmMS6^{L76H} was a malfunctioned protein and responsible for the aberrant male development of *ms6* mutant.

Additionally, expression of *AtTDF1p*-driven AtTDF1 ^{L46H}, the AtTDF1 mutant corresponding to GmMS6^{L76H}, in *attdf1* displayed consistent result. The fertilities of transgenic *AtTDF1* ^{L46H}/*attdf1* plants could not be restored because all the transformants (0/16, rescued/ transformants) developed just like *attdf1*(Fig. 6a-b, 6g, 6h-i, 6n, 6s), indicating that leucine at this position (46th in AtTDF1 and 76th in GmTDF1) is crucial for the function of TDF1.

GmMS6/GmTDF1a is the major functional TDF1 in soybean

Complementation assay above exposed that GmMS6 and GmMS6L were both functional proteins, which rose up the question why *ms6* would exhibit male sterility when there is another TDF1 coding gene in the genome. To answer that, we analyzed the expression patterns of *GmMS6* and *GmMS6L* by qRT-PCR in roots, stems, leaves, young flowers, siliques, and immature seeds. *GmMS6* displayed a typical tissue-specific expression pattern, with a much higher expression level in young flowers, which is about 6-fold higher than the 2nd highest level shown in leaves (Fig. 5). We further analyzed the expression level of *GmMS6* in petals, sepals, and pistils, and found that *GmMS6* is barely expressed in these floral parts (Fig. 5). Therefore, the high expression level detected in young flowers should be contributed by the gene expression in anthers, demonstrating that *GmMS6* is an anther-specific gene, similar to the TDF1 proteins in *Arabidopsis* and rice (Zhu et al. 2008; Cai et al. 2015). On the contrary, *GmMS6L* was expressed at a super low level in all examined tissues, indicating it is likely in the process of pseudolization (Fig. 5). Different expression patterns of *GmMS6* and *GmMS6L* illustrated that GmMS6 was the major functional TDF1 in soybean anther development, and explained why mutation at *ms6* locus would lead to male sterility.

L76H does not alter the subcellular localization or transactivation activity of GmMS6

The subcellular localization of GmMS6, GmMS6^{L76H} and GmMS6L were analyzed by transiently expressing their GFP fusion proteins driven by *35S* promoter in *N. benthamiana* leaves. Free GFP was also expressed as a control, which showed signals all over the cells (Fig. 4a). Comparatively, the fluorescence from GmMS6-GFP and GmMS6L-GFP were restricted in nucleus, the general subcellular

distribution of transcription factors (Fig. 4a). Additionally, GmMS6^{L76H}-GFP was also present in nucleus, showing that L76H would not vary the subcellular localization of GmMS6.

TDF1 is a transcriptional activator in *Arabidopsis* and rice, therefore, GmMS6 that could compensate AtTDF1's function should possess transcription activation activity as well. We then performed a transactivation activity test in yeast *AH109* strain (Fig. 4b). Yeast clones expressing GAL4 DNA binding domain (BD) could only grow on SD medium lack of Trp (SD/-Trp) but not the selective medium (SD/-Trp-His) due to no transactivation activity in BD region. Similar phenomenon was observed for yeast strain expressing BD fused GmMS6^{DBD} (BD-GmMS6^{DBD}), which was a truncated form of GmMS6 only containing the N-terminal 191 residues (the DNA binding domain, 39–146 aa). In contrast, yeast clones expressing BD-GmMS6, BD-GmMS6L and BD-GmMS6^{L76H} could grow well on both SD/-Trp and SD/-Trp-His medium, showing that these three proteins all possessed transactivation activity. Therefore, L76H doesn't affect the transactivation activity of GmMS6. As L76 is a conserved residue in R2 motif of the DNA binding domain (Fig. 2d and 2f) and L76H has no effects on protein's subcellular location or the transactivation activity, we suspected that L76H mutation likely disrupted GmMS6's function by altering its DNA binding capacity.

The DYT1-TDF1-AMS-MYB80/MYB103/MS188-MS1 genetic pathway is present in soybean

In *Arabidopsis*, TDF1 is the critical component in a well-characterized genetic pathway regulating the tapetal development and pollen wall formation, which is composed of five transcription factors, including two basic helix-loop-helix (bHLH) factors DYSFUNCTIONAL TAPETUM 1 (DYT1) and ABORTED MICROSPORES (AMS), two MYB factors TDF1 and MYB80/MYB103/MS188, and one PHD-finger protein MALE STERILITY 1 (MS1) (Zhu et al. 2011; Lu et al. 2020). In this pathway, DYT1 directly activates the expression of *TDF1*, and TDF1 subsequently promotes the expression of *AMS*. Then, AMS is required for the expression of the gene encoding MYB80/MYB103/MS188, which is an activator critical for expressing *MS1* (Fig. 7a). Depletion of any member in this cascade would lead to distorted tapetum and aborted pollens (Wilson et al. 2001; Sorensen et al. 2003; Zhang et al. 2006; Zhang et al. 2007). The same regulatory cascade known as UDT1-TDF1-TDR-OsMS188-PTC1 pathway was also identified in rice as well (Cai et al. 2015).

Soybean *ms6* exhibited similar cytological abnormities to the null mutants *attdf1* and *ostdf1*, such as vacuolated tapetum cells, undissolved callose, and crushed microspores (Fig. 1f-h; Ilarslan et al. 1999; Zhu et al. 2008; Cai et al. 2015). *AtTDF1p*-driven *GmMS6*/*GmTDF1a* was able to recover the fertility of *attdf1* mutant like *AtTDF1p*-driven *OsTDF1* did (Fig. 4d, 4k; Cai et al. 2015). These showed that TDF1's function was conserved in *Arabidopsis*, rice, and soybean, and implied that DYT1-TDF1-AMS-MYB80/MYB103/MS188-MS1 genetic pathway was likely present in soybean as well. Therefore, we performed a homology search of these transcription factors in soybean genome, and found that the whole pathway was indeed present in soybean. Moreover, all the members in this cascade had multiple paralogs, particularly, four for DYT1 and two for the others (Fig. 7a). Their expression levels in WT and

ms6 young flowers were assessed by qRT-PCR analysis. As a result, the expression level of *GmDYT1*(*Glyma.13g250200*) increased ~ 40% in *ms6* (Fig. 7b), which is consistent to that TDF1 negatively feedback-regulates the expression of *DYT1* in *Arabidopsis* (Cai et al. 2015). However, the expression levels of the other three *GmDYT1*s were not changed significantly in *ms6*, indicating the functions of DYT1s are diverged. Similarity, the expressions of two *GmTDF1* were not affected in *ms6* (Fig. 7b), suggesting that GmTDF1a is not involved in its own expression. On the other hand, the expressions of genes encoding the downstream transcription factors, like AMS, MYB80/MYB103/MS188, and MS1, were mostly downregulated in *ms6* compared to WT (Fig. 7b), which is similar to the reports in *attdf1* and *ostdf1* (Zhu et al. 2008; Cai et al. 2015). One exception was *Glyma.01g047400*, encoding a MS1 homolog, which expressed similarly in WT and *ms6* (Fig. 7b). This suggests that the expression of *Glyma.01g047400* is not controlled through the DYT1-TDF1-AMS-MYB80/MYB103/MS188-MS1 genetic pathway and implies that *Glyma.01g047400* is functionally diverged from its paralog *Glyma.02g107600* or even might lost the function in anther development. Noticeably, *Glyma.02g107600* is located at Gm02:10270908...10267934 in the region downstream of Satt157 (Gm02:9240028), which is rich of fertility controlling candidate genes, including *ms3*, *msMOS*, and *female-partial sterile-1* (*Fsp1*) (Cervantes-Martinez et al. 2009).

Additionally, we examined the expressions of two enzyme coding genes that are regulated by this transcription factor cascade and related to microspores releasing and pollen wall formation, that is, *A6* and *MALE STERILE 2* (*MS2*) (Zhang et al. 2007; Zhu et al. 2008; Wang et al. 2018). A6 protein is proposed as a member of callase complex required to degrade the callose encompassing tetrads, as its protein sequence is highly similar to β-1,3-glucanase and its gene expression is correlated to callase synthesis temporally and spatially (Hird et al. 1993). The results showed that *A6* gene only had one copy in soybean genome (Fig. 7a) and was barely expressed in *ms6* (Fig. 7b), explaining why the callose surrounding tetrads was not dissolved in *ms6* mutant. *MS2* encodes a fatty acyl reductase catalyzing the palmitoyl acyl-carrier protein into a fatty alcohol, a precursor of sporopollenin (Aarts et al. 1997; Chen et al. 2011). These precursors are subsequently transported from tapetum cells into locule and deposited onto the microspore surface to produce the sexine layer (Wang et al. 2018). *MS2* has two paralogs in soybean genome (Fig. 7a), both downregulated drastically in *ms6* (Fig. 7b). These results showed that the DYT1-TDF1-AMS-MYB80/MYB103/MS188-MS1 genetic pathway is present and functions conservatively in soybean.

Discussion

Plant male sterile mutants are important materials for studying the anther development mechanisms and crucial tools for crop hybrid breeding. So far, 13 genetic loci in soybean have been reported to condition male sterile phenotype independently when they are mutated, including *ms1-ms9*, *msp, msMOS*, *mst-M*, and ms_{NJ} (Yang et al. 2014; Zhao et al. 2019; Nie et al. 2019; Thu et al. 2019), but only *ms4* has been molecularly identified, which encodes a PHD-finger protein and is involved in the meiosis process of microsporocyte (Thu et al., 2019). Amongst these mutants, two *ms6* mutants identified decades ago

exhibit no-pollen phenotypes (Fig. 1c-d; Skorupska and Palmer 1989; Ilarslan et al. 1999), which makes them ideal genetic materials for soybean improvement by facilitating the canonical recurrent selection (Lewers et al. 1996) or the novel GMS-based hybrid-seed production technology (SPT) (Perez-Prat and van Lookeren Campagne 2002; Weber et al. 2009). Identifying the *ms6* gene is helpful to its application in recurrent selection and critical to its application in SPT development.

In the present study, we revealed that *ms6* is correlated to the mutation at the *Glyma.13g066600* locus (*GmMS6*), which encodes a TDF1 homolog (GmMS6/GmTDF1a), an R2R3 MYB transcription factor specifically expressed in anther and required for appropriate tapetum development (Fig. 2, 3, 4, and 5). The *ms6* allele present in T295H is caused by a point mutation, which leads to the substitution of L76 to H, a conserved residue on the R2 DNA-binding motif of GmMS6/GmTDF1a, suggesting that L76H likely alters DNA binding activity to disrupt the protein's function (Fig. 2). We also found that the transactivation activity and subcellular distribution of $GmMS6^{L76H}$ were not disturbed (Fig. 6) and the mutant gene in *ms6* was expressed at the WT level (Fig. 7), both supporting the above assumption from the other side. Phylogenetic and complementation analyses showed that GmMS6/GmTDF1a has a recently diverged and functional paralog GmMS6L/GmTDF1b, but the *GmMS6L/GmTDF1b* gene is expressed constitutively at a low level so that it cannot compensate the defective of *GmMS6/GmTDF1a* (Fig. 3; Fig. 5).

TDF1 is conservatively present in angiosperm species (Fig. 3), regulating the tapetal and microspore development (Zhu et al. 2008; Cai et al. 2015). One major genetic pathway it functions in *Arabidopsis* is the ordered transcription factor cascade DYT1-TDF1-AMS-MYB80/MYB103/MS188-MS1 (Zhu et al. 2011; Lu et al. 2020), which is also identified as UDT1-TDF1-TDR-OsMS188-PTC1 pathway in rice (Cai et al. 2015). It is proposed that DYT1, TDF1 and AMS are important for the early tapetum development whilst MS188 and MS1 are required for late tapetum development and pollen wall formation based on the cytological aberrations in corresponding mutants and the temporal and spatial expression pattern of each gene revealed by *in situ* hybridization (Zhu et al. 2011; Lu et al. 2020). This pathway is also critical for activating callose degeneration genes such as *A6* (Zhang et al. 2007; Zhu et al. 2008). Expressions of *OsTDF1* and *GmTDF1s* in *Arabidopsis attdf1* mutant under the control of native *AtTDF1* promoter are able to recover the fertility of the mutant plants, evidencing that TDF1's major functions are quite conserved (Fig. 4; Cai et al. 2015). However, slight divergences of *TDF1* are noticed in different species. For example, *in situ* hybridization showed that *AtTDF1* was expressed strongly and equivalently in tapetum and meiocytes at stage 6 whilst *OsTDF1* was expressed much stronger in tapetum than in meiocytes at similar development stage (Zhu et al. 2008; Zhu et al. 2011; Cai et al. 2015). Expressing *OsTDF1* in *attdf1* only partially recovered the expression levels of downstream target genes like *AMS*, *MYB80/MYB103/MS188*, and *MS1* (Zhu et al. 2008; Zhu et al. 2011; Cai et al. 2015). The function of *TDF1* in soybean is likely more diverged. Compared to *Arabidopsis* and rice, soybean possesses an extra anther wall layer between tapetum and middle layer, termed as parietal layer (Fig. 1; Ilarslan et al. 1999). In *ms6*, the parietal layer is also vacuolated and obsessively enlarged, indicating that GmMS6/GmTDF1a plays an important role in regulating parietal layer's development progress. Moreover, mutant *attdf1* and

ostdf1 can process meiosis successfully to generate tetrads, whilst both *ms6* mutants (*Ames1* and *Ames2*) showed aberrations in cytokinesis following telophase II, resulting in partially- or non-separated multi-nucleic microspores (Fig. 1f; Skorupska and Palmer 1989; Ilarslan et al. 1999).

Compared to the major crops rice and maize with dozens *ms* mutants, many of which have been cloned and well characterized (Guo and Liu 2012; Wan et al. 2019), soybean only has 13 *ms* loci reported (Yang et al. 2014; Zhao et al. 2019; Nie et al. 2019; Thu et al. 2019). One major reason should be that soybean is a paleopolyploid with two recent rounds of whole genome duplication (WGD), occurring ~ 13 and 59 million years ago, and about 75% of the genes exist with multiple copies (Schmutz et al., 2010). For example, amongst the genes we investigated in the present study, including the DYT1-TDF1-AMS-MYB103-MS1 pathway and two downstream target genes, all but *A6* have ≥ 2 paralogs in the nuclear genome (Fig. 7). Therefore, it is a big chance that nature spontaneous mutation at one microsporogenesis-related gene would not affect anther development due to another functional redundant paralog(s) in the genome. However, angiosperm genomes usually undergo diploidization soon after WGD and tend to retain only a single functional copy for most duplicated genes, and the alternative gene copies might be lost, silencing, or evolving new functions (Lynch and John 2000; Soltis et al., 2015). Similar process is ongoing for the soybean genome evolution. *GmMS6L* coding for functional protein but constantly with a minimum expression level is likely a sign of gene silencing. Another identified male sterile gene in soybean, *MS4*, also have a nonfunctional homologous copy, *MS4_homolog*, which is transcribed at a low level and codes for dysfunction protein (Thu et al., 2019). Furthermore, two paralogs of a member in TDF1 regulatory pathway, MS1, exhibited different expression in *ms6*. Only the expression of one copy, *Glyma.02g107600*, is suppressed in *ms6* while the other copy, *Glyma.01g047400*, is expressed similarly in WT and *ms6* flora tissue, suggesting that the latter gene is no longer involved in the conserved DYT1-TDF1-AMS-MYB80/MYB103/MS188-MS1 pathway. Further study is needed to assess whether it become a non-functional gene or develop some new functions.

Anther development mechanism researches in model plants like *Arabidopsis*, rice and maize have achieved significant progresses over the past decades, which have shed light on the regular genetic basis regulating angiosperm anther development (Guo and Liu 2012; Wan et al. 2019). However, soybean anther develops a distinct morphological characteristic with the parietal layer in anther wall and possesses a more complexed regulatory network due to a large extent of gene duplication. Therefore, conserved transcription factors, like GmMS6 (GmTDF1a), may evolve some new regulatory pathways in anther development process. Future investigation in GmMS6 downstream network by using RNA-Seq and ChIP-seq technologies can help to further dissect its function and enrich our understanding and knowledge in soybean anther development. Moreover, identification of the *ms6* gene in the present study provides an essential element in establishing the GMS-based SPT technology for soybean hybrid seed production.

Declarations

Acknowledgement

We thank U.S. National Plant Germplasm System (NPGS) for providing *ms6* (T295H, PI 533601) seeds for this project. We thank Prof. Zhongnan Yang (Shanghai Normal University, China) for generously supporting *attdf1* heterozygotes seeds.

Author Contributions Statement

Junping Yu, Min Xu, and Chunbao Zhang designed this project. Limei Zhao and Chunbao Zhang developed BC_5F_2 segregation population. Guolong Zhao, Ying Zhang, and Chunbao Zhang engaged in the BSA-Seq and mapping work. Junping Yu, Wei Li, Peng Wang, Aigen Fu, and Min Xu were involved in the phenotypic characterization, complementary experiment, expression analysis, protein localization, and yeast assays. Junping Yu, Guolong Zhao, Wei Li, Chunbao Zhang, and Min Xu wrote the paper, and all the authors revised the paper.

Conflict of Interest Statement

All authors declared with no conflicts and approved the manuscript.

Funding

This work was supported by the National Key Research and Development Program of China (2016YFD0101500 and 2016YFD0101503 to M.X), the National Natural Science Foundation of China (32000598 to J. Y; 31971969 to C. Z), Jilin Provincial Science and Technology Development Program (20190101007JH to C. Z) and Shaanxi Provincial Natural Science Foundation (2020JQ-578 to J.Y).

References

1. **Aarts MG, Hodge R, Kalantidis K, Florack D, Wilson ZA, Mulligan BJ, Stiekema WJ, Scott R, Pereira A** (1997) The Arabidopsis MALE STERILITY 2 protein shares similarity with reductases in elongation/condensation complexes. Plant J 12:615-623
2. **Bassam BJ, Caetano-Anollés G, Gresshoff PM** (1991) Fast and sensitive silver staining of DNA in polyacrylamide gels. Anal Biochem 196:80-83
3. **Cai CF, Zhu J, Lou Y, Guo ZL, Xiong SX, Wang K, Yang ZN** (2015) The functional analysis of OsTDF1 reveals a conserved genetic pathwan for tapetal development between rice and Arabidopsis. Sci Bull 60:1073–1082
4. **Cervantes-Martinez I, Sandhu D, Xu M, Ortiz-Pérez E, Kato KK, Horner HT, Palmer RG** (2009) The male sterility locus *ms3* is present in a fertility controlling gene cluster in soybean. J Hered 100:565-570
5. **Chang Z, Chen Z, Wang N, Xie G, Lu J, Yan W, Zhou J, Tang X, Deng XW** (2016) Construction of a male sterility system for hybrid rice breeding and seed production using a nuclear male sterility gene.

Proc Natl Acad Sci U S A 113:14145–14150

6. **Chen L, Lei D, Tang W, Xiao Y** (2011) Thoughts and practice on some problems about research and application of two-line hybrid rice. Rice Sci 18:79–85
7. **Chen W, Yu XH, Zhang K, Shi J, Schreiber L, Shanklin J, Zhang D** (2011) Male Sterile 2 Encodes a Plastid-localized Fatty Acyl ACP Reductase Required for Pollen Exine Development in Arabidopsis thaliana. Plant physiol 157:842-853
8. **Clough SJ, Bent AF** (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. Plant J 16:735–743
9. **Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L** (2010) MYB transcription factors in Arabidopsis. Trends Plant Sci 15:573–581
10. **Du M, Zhou K, Liu Y, Deng L, Zhang X, Lin L, Zhou M, Zhao W, Wen C, Xing J, Li CB, Li C** (2020) A biotechnology-based male sterility system for hybrid seed production in tomato. Plant J 102:1090-1100
11. **Guo JX, Liu YG**. (2012) Molecular control of male reproductive development and pollen fertility in rice. J Integr Plant Biol 54:967-978
12. **Hird DL, Worrall D, Hodge R, Smartt S, Paul W, Scott R** (1993) The anther-specific protein encoded by the Brassica napus and Arabidopsis thaliana A6 gene displays similarity to b-1,3-glucanase. Plant J 4:1023–1033
13. **Ilarslan H, Horner HT, Palmer RG** (1999) Genetics and cytology of a new male-sterile, female-fertile soybean mutant. Crop Sci 39:58-64
14. **Keim P, Olson TC, Shoemaker RC** (1988) A rapid protocol for isolating soybean DNA. Soybean Genet Newsletter 15:150–152(□□□□)
15. **Kim YJ, Zhang D** (2018) Molecular control of male fertility for crop hybrid breeding. Trends Plant Sci 23:53-65
16. **Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ.** (1987) MAPMAKER: an interactive computer package for constructing genetic linkage maps of experimental and natural populations. Genomics 1:174-181
17. **Levings CS 3rd. (1990)** The Texas cytoplasm of maize: cytoplasmic male sterility and disease susceptibility. Science 250:942-947
18. **Lewers KS, Palmer RG** (1993) Genetic linkage in soybean: linkage group 8. Soybean Genetics Newsletter 20:118-124
19. **Lewis KS, Martin SKS, Widrlechner MP, Palmer RG, Hedges BR** (1996) Hybrid Soybean Seed Production: Comparison of Three Methods. Crop Sci 36:1560-1567
20. **Lu JY, Xiong SX, Yin W, Teng XD, Lou Y, Zhu J, Zhang C, Gu JN, Wilson ZA, Yang ZN** (2020) MS1, a direct target of MS188, regulates the expression of key sporophytic pollen coat protein genes in Arabidopsis. J Exp Bot 71:4877-4889

21. **Lynch M, Conery JS**. (2000) The evolutionary fate and consequences of duplicate genes. Science 290:1151–1155
22. **Mercier R, Mézard C, Jenczewski E, Macaisne N, Grelon M.** (2015) The molecular biology of meiosis in plants. Annu Rev Plant Biol 66:297-327
23. **Nie Z, Zhao T, Liu M, Dai J, He T, Lyu D, Zhao J, Yang S, Gai J.** (2019) Molecular mapping of a novel male-sterile gene *msNJ* in soybean [*Glycine max (L.) Merr.*]. Plant Reprod 32:371–380
24. **Perez-Prat E, van Lookeren Campagne MM** (2002) Hybrid seed production and the challenge of propagating male-sterile plants. Trends Plant Sci 7:199–203
25. **Peterson R, Slovin JP, Chen C** (2010) A simplified method for differential staining of aborted and non-aborted pollen grains. Int J Plant Biol 1:e13
26. **Plamer RG, Gai J, Sun H, Burton JW** (2001) Production and evaluation of hybrid soybean. In J. Janick (Ed.). Plant breeding reviews (Vol.21, pp. 263–307). New York, NY: John Wiley & Sons, Inc
27. **Sanders PM, Bui AQ, Weterings K, McIntire KN, Hsu YC, Lee PY, Truong MT, Beals TP, Goldberg RB** (1999) Anther developmental defects in Arabidopsis thaliana male-sterile mutants. Sex Plant Reprod 11: 297–322
28. **Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA** (2010) Genome sequence of the paleopolyploid soybean. Nature 463:178-183
29. **Skorupska H, Palmer RG** (1989) Genetics and cytology of the ms6 male sterile soybean. J Hered 80:304–310
30. **Soltis PS, Marchant DB, Van de Peer Y, and Soltis DE** (2015) Polyploidy and genome evolution in plants. Curr Opin Genet Dev 35:119–125
31. **Song S, Wang T, Li Y, Hu J, Kan R, Qiu M, Deng Y, Liu P, Zhang L, Dong H, Li C, Yu D, Li X, Yuan D, Yuan L, Li L** (2020) A novel strategy for creating a new system of third-generation hybrid rice technology using a cytoplasmic sterility gene and a genic male-sterile gene. Plant Biotechnol J doi: 10.1111/pbi.13457. Epub ahead of print. PMID: 32741081
32. **Sorensen A, Krober S, Unte US, Huijser P, Dekker K, Saedler H.** (2003) The Arabidopsis ABORTED MICROSPORES (AMS) gene encodes a MYC class transcription factor. Plant J 33:413–423
33. **Takagi H, Abe A, Yoshida K, Kosugi S, Natsume S, Mitsuoka C, Uemura A, Utsushi H, Tamiru M, Takuno S, Innan H, Cano LM, Kamoun S, Terauchi R.** (2013) QTL-seq: rapid mapping of quantitative trait loci in rice by whole genome resequencing of DNA from two bulked populations. Plant J 74:174–183
34. **Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S** (2013) MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30:2725–2729

35. **Thu SW, Rai KM, Sandhu D, Rajangam A, Balasubramanian VK, Palmer RG, Mendu V** (2019) Mutation in a PHD-finger protein MS4 causes male sterility in soybean. BMC Plant Biol 19:378
36. **Wang K, Guo ZL, Zhou WT, Zhang C, Zhang ZY, Lou Y, Xiong SX, Yao XZ, Fan JJ, Zhu J, Yang ZN** (2018) The Regulation of Sporopollenin Biosynthesis Genes for Rapid Pollen Wall Formation. Plant Physiol 178:283-294
37. **Wan X, Wu S, Li Z, Dong Z, An X, Ma B, Tian Y, Li J** (2019) Maize Genic Male-Sterility Genes and Their Applications in Hybrid Breeding: Progress and Perspectives. Mol Plant 12:321-342
38. **Weber N, Commuri P, Rood T and Townsend R** (2009) Petition for the Determination of Nonregulated Status for Maize 32138 SPT Maintainer Used in the Pioneer Seed Production Technology (SPT) Process. Submitted to the USDA-APHIS by Pioneer Hi-Bred International, Inc. Available at: http://www.aphis.usda.gov/brs/aphisdocs/08_33801p.pdf (last accessed 13 August 2015).
39. **Wilson ZA, Morroll SM, Dawson J, Swarup R, Tighe PJ.** (2001) The Arabidopsis MALE STERILITY 1 (MS1) gene is atranscriptional regulator of male gametogenesis, with homology to the PHD-finger family of transcription factors. Plant J 28:27–39
40. **Wu Y, Fox TW, Trimnell MR, Wang L, Xu RJ, Cigan AM, Huffman GA, Garnaat CW, Hershey H, Albertsen MC.** (2016) Development of a novel recessive genetic male sterility system for hybrid seed production in maize and other cross-pollinating crops. Plant Biotechnol J 14:1046-1054
41. **Yang Y, Speth BD, Boonyoo N, Baumert E, Atkinson TR, Palmer RG, Sandhu D** (2014) Molecular mapping of three male-sterile, female-fertile mutants and generation of a comprehensive map of all known male sterility genes in soybean. Genome 57:155-160
42. **Zhang D, Luo X, Zhu L** (2011) Cytological analysis and genetic control of rice anther development. J Genet Genomics 38:379-390
43. **Zhang D, Chang E, Yu X, Chen Y, Yang Q, Cao Y, Li X, Wang Y, Fu A, Xu M** (2018a) Molecular Characterization of Magnesium Chelatase in Soybean [Glycine max (L.) Merr.]. Front Plant Sci 9:720
44. **Zhang D, Wu S, An X, Xie K, Dong Z, Zhou Y, Xu L, Fang W, Liu S, Liu S, Zhu T, Li J, Rao L, Zhao J, Wan X** (2018b) Construction of a multicontrol sterility system for a maize male-sterile line and hybrid seed production based on the ZmMs7 gene encoding a PHD-finger transcription factor. Plant Biotechnol J 16:459-471
45. **Zhang W, Sun Y, Timofejeva L, Chen C, Grossniklaus U, Ma H** (2006) Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development* 133:3085–3095
46. **Zhang ZB, Zhu J, Gao JF, Wang C, Li H, Li H, Zhang HQ, Zhang S, Wang DM, Wang QX, Huang H, Xia HJ, Yang ZN** (2007) Transcription factor AtMYB103 is required for anther development by regulating tapetum development, callose dissolution and exine formation in Arabidopsis. Plant J 52:528–538
47. **Zhao L, Sun H, Wang S, Wang Y, Huang M, Li J** (2004) Breeding of hybrid soybean HybSoy 1. Chinese Journal of Oil Crop Sciences 26:15-17 (in Chinese)
48. **Zhao Q, Tong Y, Yang C, Yang Y, Zhang M.** (2019) Identification and mapping of a new soybean Male-Sterile Gene, *mst-M.* Front Plant Sci 10:94

49. **Zhu J, Chen H, Li H, Gao J, Jiang H, Wang C, Guan Y, Yang Z** (2008) Defective in Tapetal Development and Function 1 is essential for anther development and tapetal function for microspore maturation in Arabidopsis. Plant J 55:266–277
50. **Zhu J, Lou Y, Xu X, Yang ZN.** (2011) A genetic pathway for tapetum development and function in Arabidopsis. J Integr Plant Biol 53:892-900

Figures

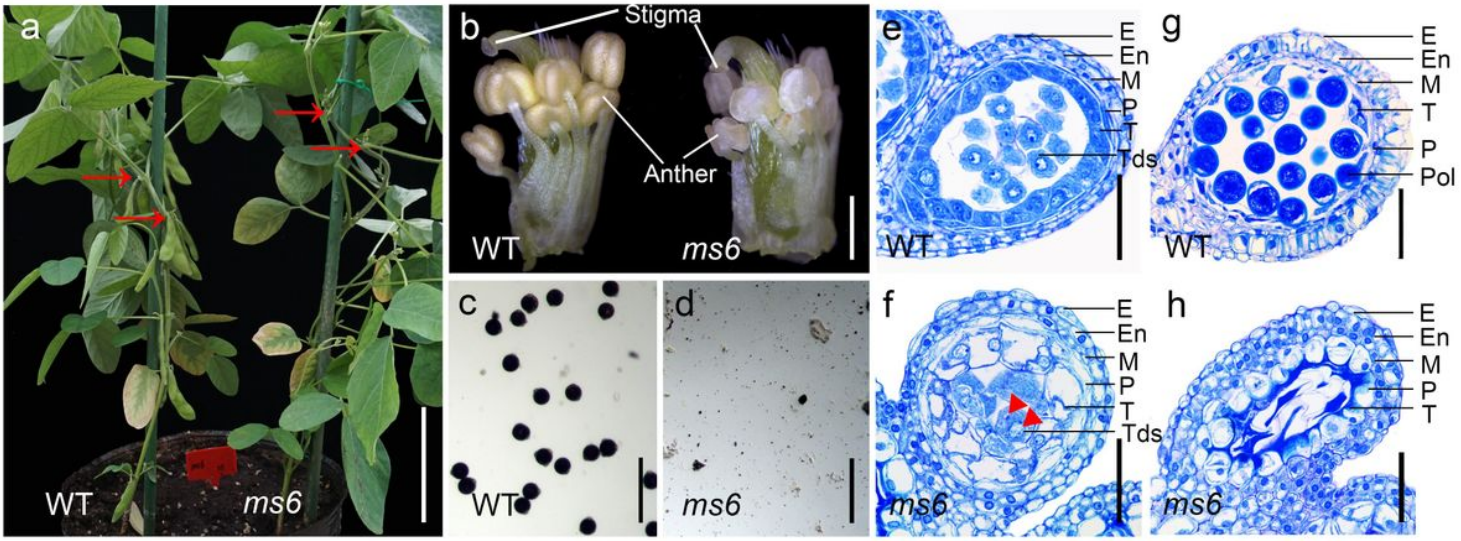

Figure 1

Phenotypic characterization of ms6 in soybean (a) Wild-type (WT) and ms6 plants at R3 Stage. Compared with WT, ms6 fails to produce elongated pods at nodes (red arrow). Scale bar = 10 cm. (b) WT and ms6 flowers with petals and sepals peeled off. Scale bar = 0.5 mm. (c, d) I2-KI staining of pollens squeezed out from WT and ms6 anthers. Scale bar = 50 μm. (e-h) Semi-thin sections of WT and ms6 anther lobes at late tetrads stage (e, f) and late microspore stage (g, h). Scale bar = 50 μm. Red triangles indicate the multi-nucleuses in the tetrads. E, Epidermis layer; En, Endodermis layer; M, Middle layer; P, Particle layer; T, Tapetal cell layer; Tds, Tetrads; Pol, Pollen.

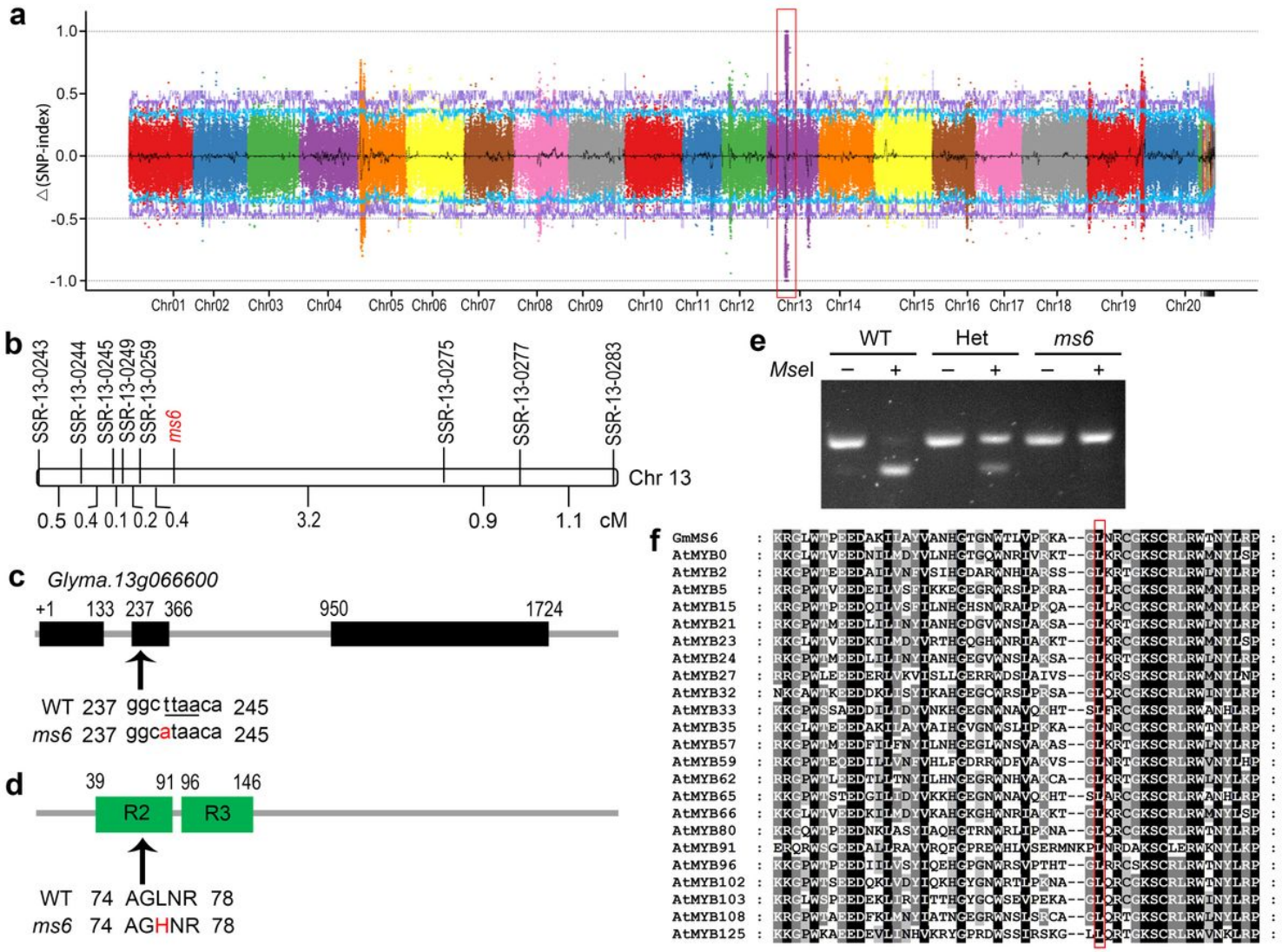

Figure 2

GmMS6 encodes a R2R3 MYB transcription factor (a) Manhattan plot of the SNP-index differences, Δ(SNP-index), between homozygous WT and ms6 bulks constructed with BC5F2 mapping population derived from ms6 x JiuB. The red box highlights the region co-segregated with ms6. (b) Genetic linkage map of ms6 on the chromosome 13. Genetic distances between two adjacent loci are indicated by centimorgans (cM). (c) GmMS6 gene structures and the mutated site in ms6. Black box represents the exon of GmMS6. Arrow indicates the SNP location and red nucleotide is the mutated site in ms6. Underlined sequence is the MseI restriction site. (d) GmMS6 protein structure. Green blocks indicate the R2R3 MYB DNA-binding domain. Arrow and red amino acid indicate the amino acid substitution site in ms6. (e) Genotyping the plants with the CAPS marker based on the SNP between WT and ms6. The MseI-undigested (-) or digested (+) PCR products are segregated on 4% agarose gel. WT, homozygous wild-type plant; Het, heterozygous (Ms6ms6); ms6, homozygous ms6 mutant. (f) Multiple sequence alignment of R2 motifs from different MYB proteins. The residue corresponding to L76 in GmMS6 (red box indicated) is conserved.

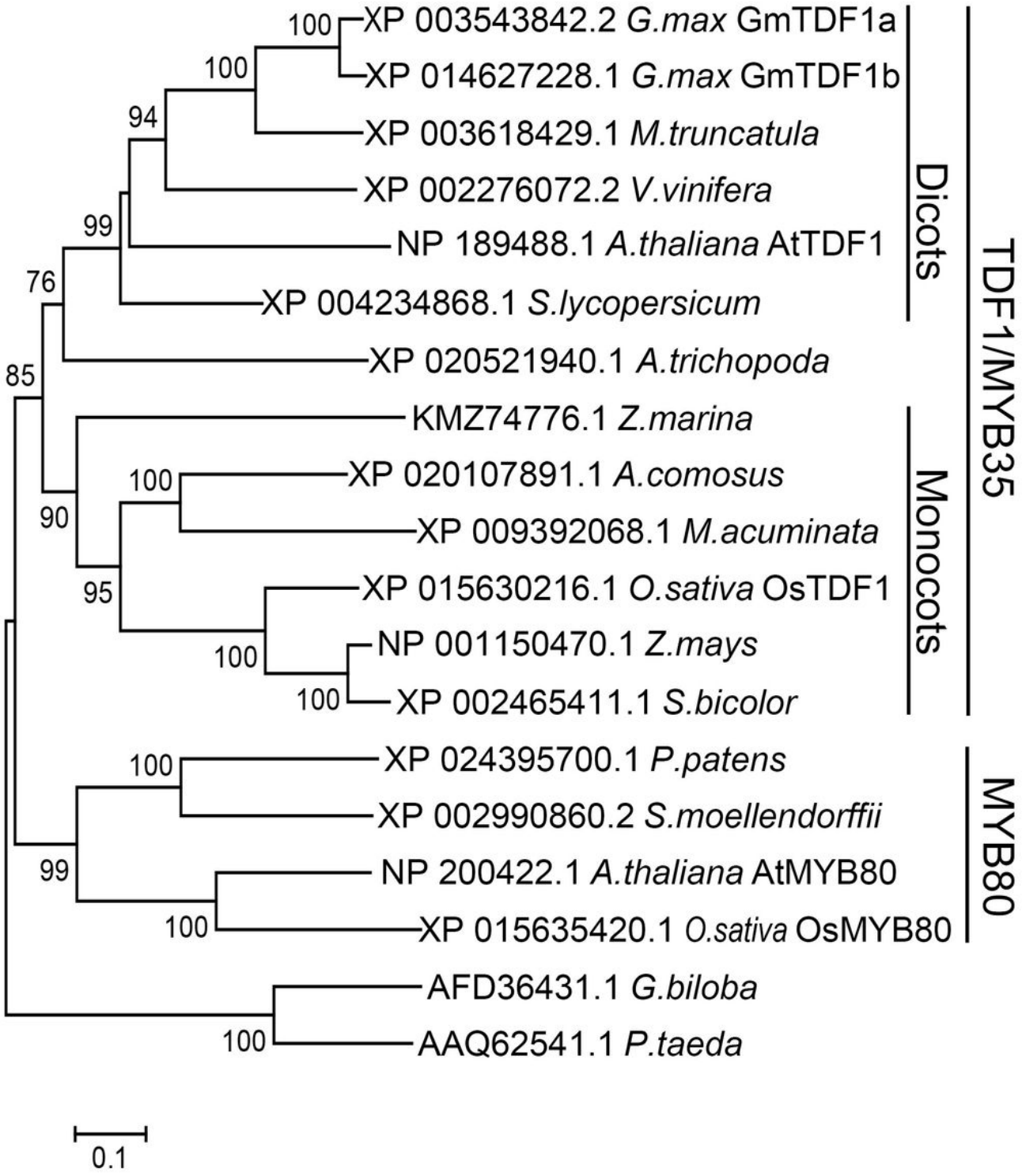

Figure 3

GmMS6 is conserved in both dicot and monocot but not in gymnosperm An unrooted maximum likelihood NJ phylogenetic tree based on GmMS6 related sequences is created by MEGA6. A. thaliana (Arabidopsis thaliana), A. trichopoda (Amborella trichopoda), A. comosus (Ananas comosus), G. biloba (Ginkgo biloba), G. max (Glycine max), M. acuminata (Musa acuminata), M. truncatula (Medicago truncatula), O. sativa (Oryza sativa), P. taeda (Pinus taeda), P. patens (Physcomitrella patens), S. bicolor

(Sorghum bicolor), S. lycopersicum (Solanum lycopersicum), S. moellendorffii (Selaginella moellendorffii), V. vinifera (Vitis vinifera), Z. mays (Zea mays) and Z. marina (Zostera marina). Bootstrap values are indicated as percentages.

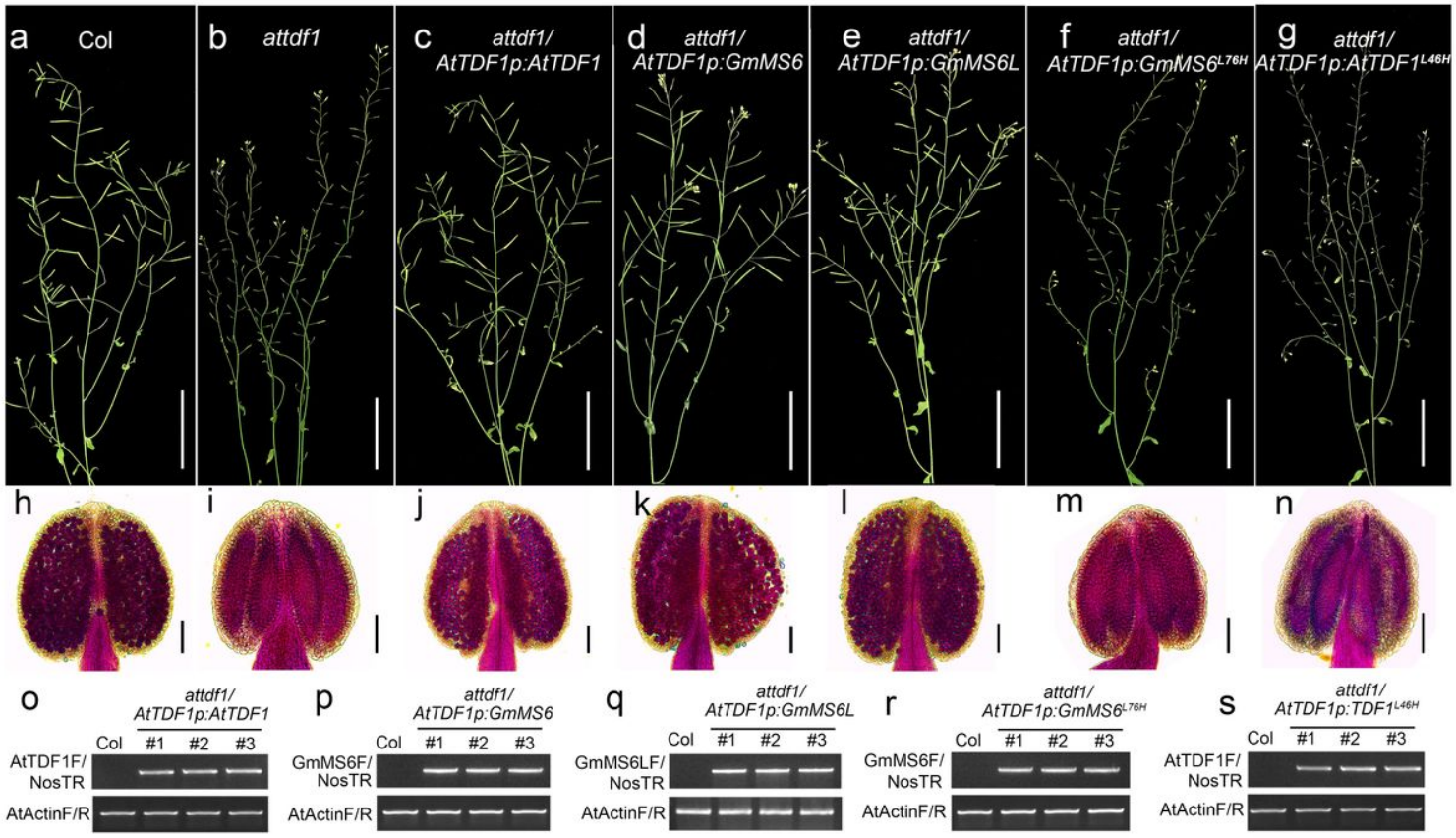

Figure 4

GmMS6 can fully rescue the attdf1 fertility (a-g) Seed-setting staged plants of col, attdf1, and transgenic positive transformants. Scale bar = 15 cm. (h-n) Alexander staining of the pollens from col, attdf1 and corresponding transgenic plants placed above them. Scale bar = 100 μm. (o-s) RT-PCR analysis shows that transgenes are expressed in the flowers of transgenic plants. AtActin are used as the native control to indicate the RNA extraction and reverse transcription process are correct. AtTDF1F, GmMS6F, and GmMS6LF are forward primers specific to the target coding sequences and NosTR is reverse primer specific to Nos terminator sequence on the vector.

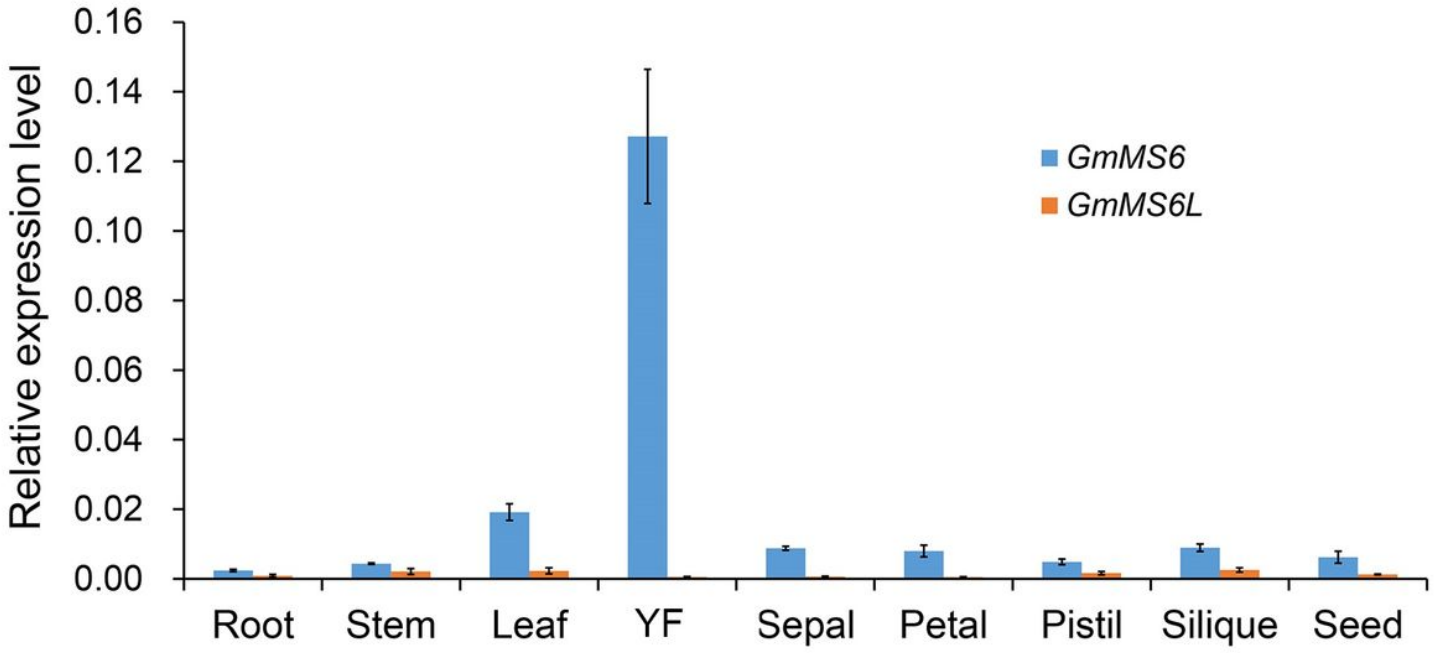

Figure 5

The relative expression patterns of GmMS6 and GmMS6L in different tissues Samples are repeated three times. The error bars denote ± SE. YF, Young Flowers.

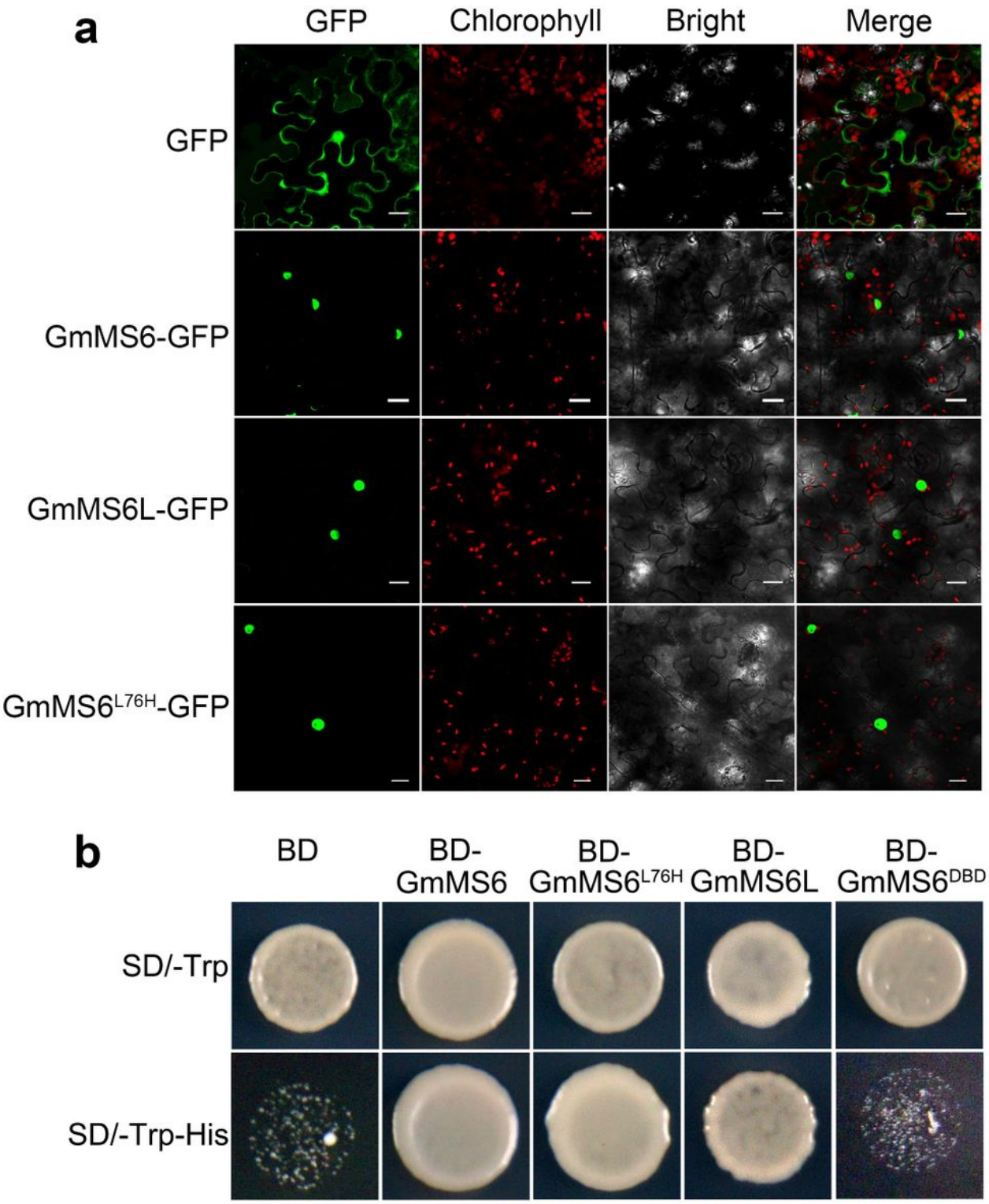

Figure 6

GmMS6 is located in the nucleus and possesses transcriptional activation activity (a) Transient expression of GFP, GmMS6-GFP, GmMS6L-GFP, and GmMS6L76H-GFP in N. benthamiana. For each protein, the images of GFP, chloroplast auto-fluorescence, bright field, and merged signals are presented. Scale bar = 20 μm. (b) Transactivation activity assay in yeast. The yeast strain transformed with empty BD vector (BD, negative control) and BD- GmMS6DBD (1-191 aa of GmMS6) without transactivation

activity can grow on synthetic dropout medium SD/-Trp but not on SD/-Trp-His. Yeast strain transformed with BD-fused GmMS6, GmMS6L76H, and GmMS6L can grow on both types of media, showing they have transactivation activity.

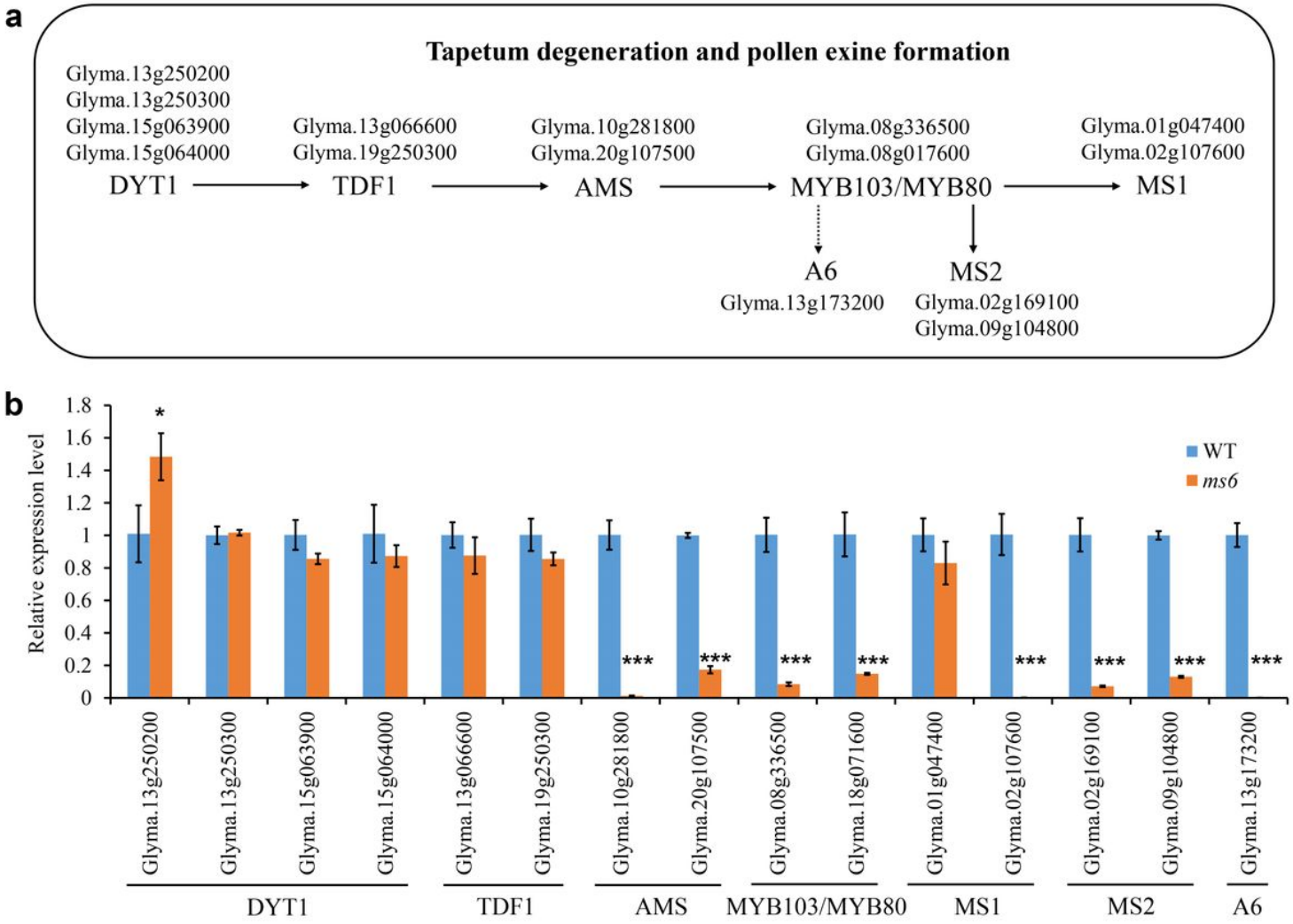

Figure 7

Relative expression level of anther development core regulators in WT and ms6 (a) Flow chart of known core factors identified in tapetum function and pollen exine formation in Arabidopsis and their orthologous in Glycine max. Solid lines represent the direct activation, and dotted lines indicate the unidentified downstream target. (b) Relative expression level of the annotated anther development key factors in WT and ms6. Samples are repeated three times. The error bars denote ± SE. P-values are calculated by using students t-test and significance levels are represented as follows: *P values < 0.05, **P values < 0.01, ***P values < 0.001.

Supplementary Files

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

- MS6TAG20210121supplementary.docx