Unusual Case of Apparent Hypermutation in *Arabidopsis thaliana*

Taku Sasaki, Ulf Naumann, Petar Forai, Antonius J. M. Matzke,[1] and Marjori Matzke[1,2]
Gregor Mendel Institute of Molecular Plant Biology, Austrian Academy of Sciences, A-1030 Vienna, Austria

ABSTRACT The *dms4* (*defective in meristem silencing 4*) mutant of *Arabidopsis thaliana* is unique in having defects in both RNA-directed DNA methylation (RdDM) and plant development. DMS4 is an evolutionarily conserved, putative transcription factor of the Iwr1 (interacts with RNA polymerase II) type. DMS4 interacts with Pol II and also with RNA polymerases IV and V, which function in RdDM. Interactions with multiple polymerases may account for the diverse phenotypic effects of *dms4* mutations. To dissect further the roles of DMS4 in RdDM and development, we performed a genetic suppressor screen using the *dms4-1* allele, which contains in the sixth intron a splice site acceptor mutation that alters splicing and destroys the open reading frame. Following mutagenesis of *dms4-1* seeds using ethyl methanesulfonate (EMS), we retrieved four dominant intragenic suppressor mutations that restored DMS4 function and wild-type phenotypes. Three of the four intragenic suppressor mutations created new splice site acceptors, which resulted in reestablishment of the wild-type open reading frame. Remarkably, the intragenic suppressor mutations were recovered at frequencies ranging from 35 to 150 times higher than expected for standard EMS mutagenesis in *Arabidopsis*. Whole-genome sequencing did not reveal an elevated mutation frequency genome-wide, indicating that the apparent hypermutation was confined to four specific sites in the *dms4* gene. The localized high mutation frequency correlated with restoration of DMS4 function implies an efficient mechanism for targeted mutagenesis or selection of more fit revertant cells in the shoot apical meristem, thereby rapidly restoring a wild-type phenotype that is transmitted to future generations.

RNA-directed DNA methylation (RdDM) is a small RNA-mediated epigenetic modification that contributes to transcriptional silencing of transposons and repetitive sequences in plants. Forward genetic screens have retrieved a number of specialized factors required for RdDM, including subunits of atypical RNA polymerases, called Pol IV and Pol V, as well as chromatin remodelers, transcription factors, RNA binding proteins, and other novel factors whose precise functions in the RdDM pathway are unclear (Matzke *et al.* 2009; Haag and Pikaard 2011). Despite the presumed role of RdDM in transposon silencing, most mutants defective in this process do not mobilize transposons nor do they display overt developmental phenotypes. An exception is *dms4* (*defective in meristem silencing 4*), which is the only mutant identified so far that is impaired in both RdDM and plant development (He *et al.* 2009; Kanno *et al.* 2010).

We identified *dms4* mutations in a forward genetic screen for mutants deficient in RdDM and transcriptional gene silencing of a *GFP* reporter gene in shoot and root meristem regions in *Arabidopsis thaliana* (*Arabidopsis*) (Kanno *et al.* 2008, 2010). In addition to reducing RdDM and releasing *GFP* silencing, mutations in *DMS4* condition a pleiotropic developmental phenotype characterized by late seed germination, dwarf stature, pale and serrated leaves, late flowering, small flowers, and abnormal phyllotaxy (Kanno *et al.* 2010).

DMS4, identified as RDM4 in a separate genetic screen (He *et al.* 2009), is an evolutionarily conserved, putative transcription factor of the Iwr1 type (**i**nteracts **w**ith **R**NA polymerase II). Iwr1 was first identified in a global proteomics analysis in budding yeast, where it was shown to interact with

Copyright © 2012 by the Genetics Society of America
doi: 10.1534/genetics.112.144634
Manuscript received August 2, 2012; accepted for publication September 18, 2012
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/content/suppl/2012/09/28/genetics.112.144634.DC1.
Sequence data from this article have been deposited with the Sequence Read Archive at NCBI under accession nos. wild-type T+S (SRX159802); *dms4-1* (SRX170675); *dms4-1s1* (6-3-5) (SRX170682); *dms4-1s1* (12-1-3) (SRX170683); *dms4-1s3* (3-2-3) (SRX170684); *dms4-1s3* (3-4-2) (SRX170685); and *dms4-1s4* (26-4a) (SRX170686).
[1]Present address: Institute of Plant and Microbial Biology, Academia Sinica, 128 Sec. 2, Academia Rd. Nankang, Taipei 11529, Taiwan.
[2]Corresponding author: Institute of Plant and Microbial Biology, Academia Sinica, 128Sec. 2, Academia Rd., Nankang, Taipei 11529, Taiwan. Email: marjorimatzke@gate.sinica.edu.tw.

many subunits of RNA polymerase II (Pol II) (Collins *et al.* 2007). Recent work in budding yeast has demonstrated that Iwr1 is required for nuclear import of Pol II (Czeko *et al.* 2011) and for transcriptional initiation by Pols I, II, and III (Esberg *et al.* 2011). In *Arabidopsis*, DMS4 has been shown to interact with Pol II (He *et al.* 2009) and also with Pol IV (Law *et al.* 2011) and Pol V (He *et al.* 2009). The diverse phenotypic effects of *dms4* mutations may thus reflect interactions with multiple RNA polymerases (He *et al.* 2009; Kanno *et al.* 2010). However, the precise function of DMS4 in various Pol II-, Pol IV-, and Pol V-dependent processes remains unknown.

To dissect the roles of DMS4 in RdDM and development, we have carried out a genetic suppressor screen in *Arabidopsis* to identify second site mutations that either restore RdDM/GFP silencing, normal development, or both. All three categories of mutant have been retrieved in this screen, demonstrating that it is possible to genetically separate the effects of a *dms4* mutation on RdDM and development. Here we report on four dominant, intragenic suppressor mutations that reestablish both RdDM and normal development by restoring the wild-type function of the DMS4 protein. Remarkably, these intragenic suppressor mutations were observed at frequencies that range from ~35 to 150 times higher than normally observed with standard EMS mutagenesis in *Arabidopsis*. We discuss this unexpectedly high mutation frequency in the context of targeted mutagenesis of the *dms4* gene or selection of more fit revertant cells in the shoot apical meristem.

Materials and Methods

Plant materials and EMS mutagenesis

The two-component transgene silencing system (T+S) used to identify the *dms4-1* mutant is described in previous publications (Kanno *et al.* 2008, 2010) and illustrated in Supporting Information, Figure S1. The *Arabidopsis* gene identifier number of *DMS4* is At2g30280. Approximately 14,400 BC_1F_3 seeds of *dms4-1* (Kanno *et al.* 2010) were mutagenized by treatment with 1% EMS (Sigma-Aldrich) for 3 hr. Mutagenized seeds (M_1 generation) were sown on soil and grown in a growth chamber at 21° under a 16-h light/8-h dark regime. Approximately 10,600 M_1 plants grew from the mutagenized seed. M_1 plants with a wild-type–like phenotype were identified, the *dms4* gene was sequenced, and self-pollinated seeds (M_2 generation) were harvested from each individual M_1 plant showing a wild-type phenotype. All other M_1 plants were pooled into 144 batches (containing ~74 M_1 plants/batch), and the M_2 seeds were harvested in bulk from each batch. Approximately 400–500 M_2 seedlings per batch (~5–7 M_2 seedlings from each M_1 plant) were sown on solid Murashige and Skoog (MS) medium and GFP^- plants were screened using a Leica fluorescence stereomicroscope. For the second EMS mutagenesis treatment, ~5500 BC_2F_4 *dms4-1* seeds were treated as described above, and M_2 seeds from ~3575 M_1 plants were harvested and screened for GFP. "BC" refers to a backcross of the *dms4* mutant to the wild type T+S line, with the subscript number indicating the number of times the backcross has been carried out. "F" refers to generations of self-fertilization following a backcross.

DNA methylation analyses

Genomic DNAs were extracted from mature rosette leaves using DNeasy Plant Mini kit (Qiagen). We analyzed methylation of endogenous sequences by Chop–PCR, an assay in which genomic DNA is digested ("chopped") with a methylation-sensitive restriction enzyme, and then used as a template for PCR amplification using primers flanking the restriction enzyme site (Earley *et al.* 2010). For this, 50 ng of genomic DNA was digested with *Hae*III, which reports on CHH methylation (H is A, T, or C), in 20 μl of reaction mix at 37° overnight. After restriction digestion, 1 μl of digested DNA was used as template for PCR in the 10-μl reaction mix. Primers used for Chop–PCR are listed in Table S1.

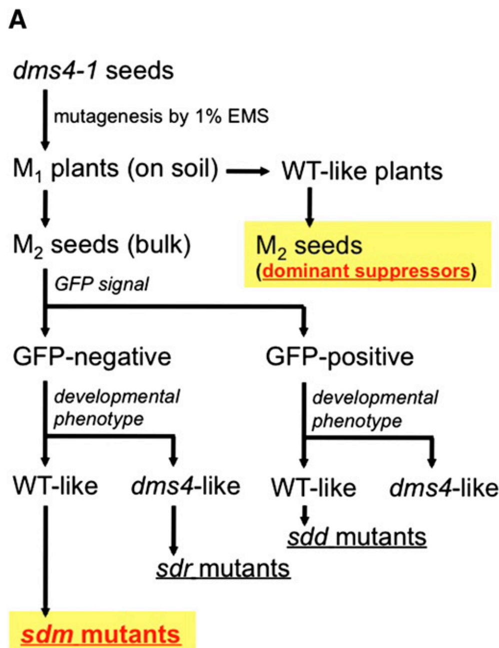
For the bisulfite sequencing analysis of the target enhancer region, 1 μg of genomic DNA, isolated from rosette leaves as described above, was digested with *Hin*dIII, and then 500 ng was used for bisulfite conversion of unmethylated cytosines to uracil using the EpiTect Bisulfite kit (Qiagen). Amplified fragments were cloned using pGEM-T Easy Vector system (Promega) and 10–20 clones were sequenced. Complete conversion was confirmed by analyzing the methylation-free *PHAVOLUTA* (*PHV*) locus (Reinders *et al.* 2008). The primers used are shown in Table S1.

Analysis of dms4 transcription

Total RNA was extracted from rosette leaves using the RNeasy Plant Mini kit (Qiagen) and treated with TURBO DNA-free DNase (Ambion). Approximately 2.5–3.0 μg of total RNA was reverse transcribed using the ReverseAid H Minus First Strand cDNA Synthesis kit (Fermentas). RT–PCR products using primer pair 3028-3 and DMS4sqR2 (Table S1) were cloned using the pGEM-T Easy Vector system and sequenced.

Western blot analysis

Nuclei were extracted from the following lines: T+S, *dms4-1*, *dms4-1s1*, *dms4-1s2*, *dms4-1s3*, and *dms4-1s4*. Around 1.5 g of frozen rosette leaves were ground in liquid nitrogen and resuspended in 30 ml extraction buffer 1 [0.4 M sucrose, 10 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 5 mM β-mercaptoethanol] supplemented with EDTA-free Proteinase Inhibitor cocktail (Roche). The suspension was centrifuged at 2,000 × *g* for 15 min at 4°. The supernatant was discarded and the pellet was resuspended in 1 ml of extraction buffer 2 [0.25 M sucrose, 10 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 1% Triton-100, 5 mM β-mercaptoethanol] supplemented with EDTA-free proteinase inhibitor, and centrifuged at 2,000 × *g* for 10 min at 4°. This step was repeated. The final pellet was resuspended in nuclei lysis buffer [50 mM Tris-HCl (pH 8.0),



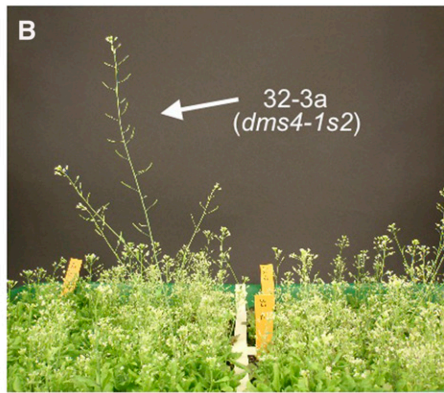


Figure 1 Flow chart of *dms4* suppressor screening and phenotype of M_1 suppressor mutant. (A) The *dms4-1* mutant is GFP^+, due to loss of RdDM at the upstream target enhancer, and displays a distinctive developmental phenotype. Seeds of the *dms4-1* mutant were treated with 1% EMS and sown on soil. Wild-type–looking (WT) M_1 plants were identified as dominant suppressors and M_2 seeds obtained from self-fertilization were harvested from individual WT-appearing plants. All other M_1 plants were pooled into batches and the M_2 seeds harvested in bulk from each batch (see details in *Materials and Methods*). M_2 seedlings were germinated on plates and screened for GFP^- and GFP^+ phenotypes, and these groups were further subdivided based on their developmental phenotype (WT-like or *dms4-1*–like). The *sdm* (**s**uppressor of ***dm**s4*) mutants are GFP^- and have a WT phenotype; *sdr* (**s**uppressor of ***d**ms4*, **R**dDM) mutants are GFP^- but have a *dms4-1*–like phenotype; and *sdd* (**s**uppressor of ***d**ms4*, **d**evelopment) mutants are GFP^+ but have a WT phenotype. In this article, we describe four *sdm* mutants that contain dominant, intragenic suppressor mutations. (B) An M_1 suppressor mutant (strain 32-3a) containing the *dms4-1s2* suppressor mutation has a WT phenotype, being much taller than the surrounding dwarf *dms4* mutant plants and displaying normal flowers and spiral phyllotaxy.

10 mM EDTA, 1% SDS] supplemented with EDTA-free Proteinase Inhibitor, and sonicated 3 × 10 sec, 40% duty cycle, and 20% power. Proteins were separated by SDS–PAGE (10% gel), transferred to a PVDF membrane (Millipore), followed by Western blotting according to standard procedures. Rabbit anti-DMS4 polyclonal antibodies were generated by Eurogentec (Belgium) using overexpressed DMS4 protein in bacteria and used at 1:1000 dilutions. Secondary antibody [goat antirabbit IgG-conjugated with horseradish peroxidase (Biorad)], was used at 1:10,000 dilution. The blots were developed using an enhanced chemoluminescence kit (AmershamPharmacia Biotech).

Whole-genome sequencing

Total DNA was extracted from 1.5 g of root cultures grown in liquid MS medium with shaking using the CTAB method (Eun *et al.* 2011). For each suppressor mutant, we isolated DNA from root cultures of approximately five pooled M_3 plants. Around 400 ng of DNA was sonicated using a Covaris S2 (Covaris, Woburn, MA) to produce fragments ~300–800 bp in length for making sequence libraries for paired-end reads.

Data analysis

Reads of whole-genome sequence were assembled to the *Arabidopsis* genomic sequence of TAIR9 (http://www.arabidopsis.org/) using CLC Genomics Workbench (CLC bio). To reduce the effect of sequencing errors, we discarded 5′- and 3′-end nucleotides of each read to remove low-quality bases. To reduce errors caused by presequencing amplification, we removed reads that had identical locations of starts and ends. Because each of the five plants used for DNA isolation was still segregating mutations induced by EMS during the suppressor screen, it was conceivable that at a minimum, only one of these plants still contained an induced mutation in the heterozygous state (*i.e.*, in 10% of the pooled DNA isolated from five diploid plants). Therefore, mismatched nucleotides with more than five reads and >10% coverage were counted as single nucleotide polymorphisms (SNPs) induced by EMS mutagenesis. We focused on G/C to A/T changes, which typically result from EMS mutagenesis (Greene *et al.* 2003), and SNPs in suppressor mutants were compared with those in the T+S and *dms4-1* lines. There were 319 common SNPs among the five suppressor mutants; however, 294 of them were already present in T+S and/or *dms4-1* lines with fewer than five reads and/or <10% coverage. Of the remaining 25 SNPs, 15 were shown by sequencing to preexist in the mutagenized *dms4-1* seed population and were therefore removed from the analysis. Of the final 10 common SNPs, 7 were within repetitive sequences, making it difficult to sequence a specific copy, and 3 were undetectable in the small population of *dms4-1* seedlings tested. Nevertheless, the simplest explanation is that these 10 SNPs preexisted in the original *dms4-1* seed population.

Results

dms4-1 suppressor screen: dominant intragenic suppressor mutations in M_1 plants

The scheme for the suppressor screen is illustrated in Figure 1A. Here we focus on mutations that restore both RdDM/

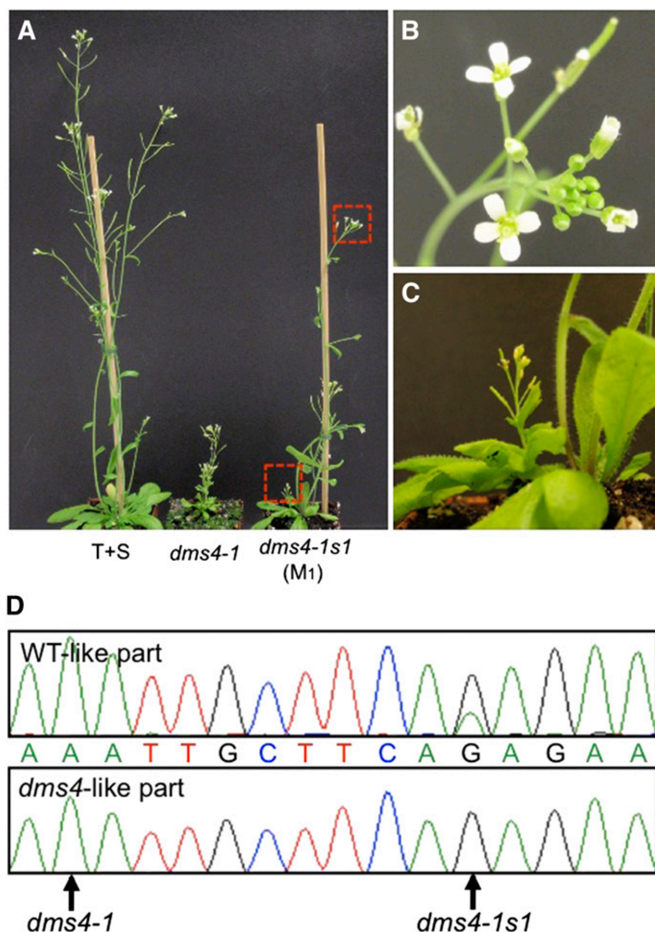


Figure 2 Chimeric M_1 suppressor plant. (A) Photographs of the wild-type (WT) T+S line, the *dms4-1* mutant, and an M_1 plant that is chimeric for a dominant intragenic suppressor mutation (*dms4-1s1*). This plant comprises a *dms4-1*–like sector (bottom red box) and a WT sector with normal flowers (top red box). Close ups of boxed regions are shown in B and C, respectively. (D) Sequencing chromatograms of the *DMS4* gene in the WT sector (top) and *dms4-1*–like sector (bottom) of the chimeric M_1 plant shown in A. Arrows indicate positions of the recessive *dms4-1* mutation, which is homozygous in both sectors, and the dominant *dms4-1s1* mutation, which is heterozygous in the WT sector.

GFP silencing and normal development (dominant *sdm* mtuants, Figure 1A). For the screen, we used the *dms4-1* allele, which has a mutation in the splice site acceptor of the sixth intron. This mutation disrupts the open reading frame, leading to an altered amino acid sequence in the final third of the DMS4 protein (Kanno *et al.* 2010). Approximately 14,400 *dms4-1* seeds (BC_1F_3 generation) were treated with EMS according to standard protocols and sown on soil. As the M_1 plants were bolting, flowering and setting seed, we noticed nineteen that had wild-type characteristics, being taller than *dms4-1* mutants and displaying normal leaves, phyllotaxy, and flowers (Figure 1B). Several of these M_1 plants appeared chimeric, containing both wild-type and *dms4-1* sectors (Figure 2, A–C).

To rule out that these wild-type-appearing M_1 plants were due to contamination by wild-type seeds, we sequenced the *dms4* gene in DNA isolated from leaves of 7 of these plants (the *dms4* gene was sequenced in the remaining 12 in the M_2 generation) (Table S2). All of the wild-type–appearing plants were homozygous for the original *dms4-1* mutation and in addition, each was heterozygous for a second mutation within the *dms4* gene. Four different G to A intragenic suppressor mutations, all within 200 bp of the original *dms4-1* mutation, were identified (Figure 3A). Chimeric plants containing wild-type and *dms4*-like sectors were found to be genetic mosaics in which the heterozygous suppressor mutation was detected only in the wild-type leaves (Figure 2D).

Because EMS-induced mutations in the M_1 generation are still heterozygous, all of the intragenic suppressor mutations that restore a wild-type phenotype act as dominant mutations. We named the suppressor mutations *dms4-1s1*, *dms4-1s2*, *dms4-1s3*, and *dms4-1s4* (Figure 3A). These four intragenic suppressor mutations were recovered multiple times in the M_1 population (Table 1). None of the suppressor mutations were observed following a mock treatment of *dms4-1* seeds, confirming they did not preexist or occur spontaneously in the mutagenized seed population. Because each M_1 plant was derived from an individually mutagenized seed, we could conclude that the suppressor mutations were induced independently multiple times during EMS mutagenesis.

All 19 of the wild-type–appearing M_1 plants containing the dominant intragenic suppressor mutations spawned variable numbers of GFP^- M_2 progeny (Table S2), indicating that *GFP* silencing was restored in plants containing a suppressor mutation (Figure 4A). Consistent transmission to the M_2 generation indicated that the *dms4-1* suppressor mutations were present in germ cell progenitors in the L2 layer of the shoot apical meristem (SAM) of M_1 plants.

M_2 plants: inheritance of intragenic suppressor mutations and restoration of GFP silencing

Screening for GFP^- seedlings in the M_2 progeny of the remaining M_1 plants (approximately five to seven M_2 seedlings per M_1 plant were sampled) identified additional cases of one of the four previously identified intragenic suppressor mutations (Table 1, Table S3). When transferred to soil and grown to maturity, all of the GFP^- M_2 plants lacked the developmental defects associated with the *dms4-1* mutation. In addition to restoring normal development and *GFP* silencing, the *dms4-1* suppressor mutations fully reestablished DNA methylation of the target enhancer driving *GFP* expression (Figure 4B) and several endogenous targets of RdDM (Figure 4C).

Because the intragenic suppressor mutations were identified in plants grown from separate batches of M_2 seeds that did not contain seeds from the suppressor mutants identified in the M_1 generation, they are believed to have arisen independently during EMS mutagenesis. In principle—given the dominant nature of the suppressor mutations—the parents of these M_2 suppressor mutants could have been detected in the M_1 generation but they were apparently

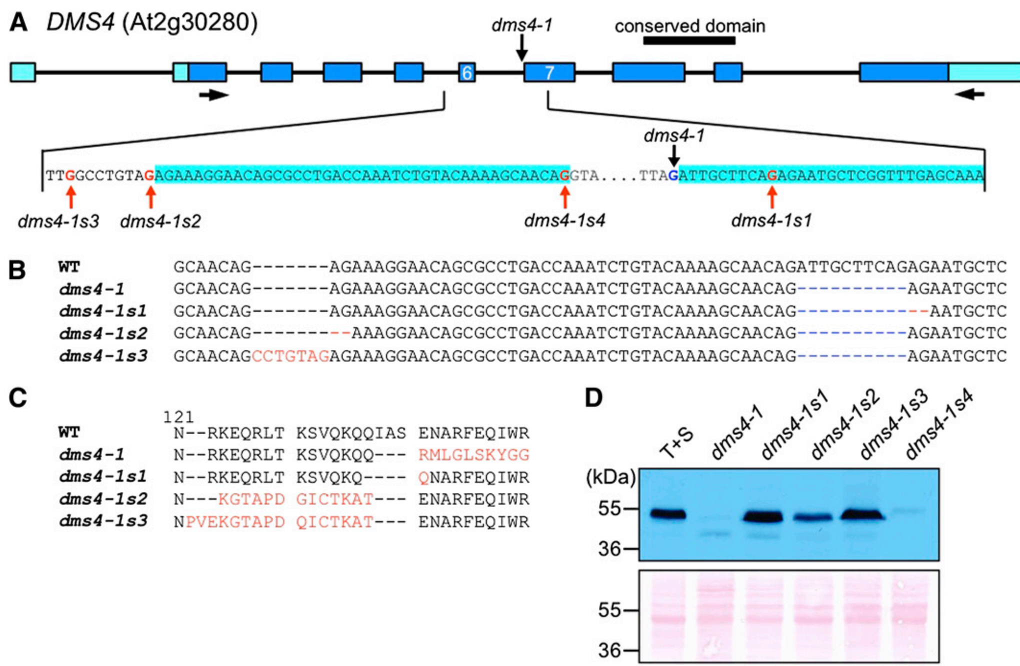


Figure 3 Positions of *dms4-1* intragenic suppressor mutations and effects on transcription and translation. (A) Schematic structure of *DMS4* gene. Boxes in light blue and blue indicate UTRs and coding regions, respectively. Horizontal arrows indicate positions of primers used for RT–PCR. Black and red vertical arrows indicate, respectively, positions of the *dms4-1* mutation and the four dominant intragenic suppressor mutations: *dms4-1s1* to *dms4-1s4*. (B) cDNA sequences of *DMS4* transcripts from the WT T+S line, *dms4-1*, and three suppressor mutants. The blue and red letters/dashes indicate changes caused by the *dms4-1* mutation and suppressor mutations, respectively. (C) Predicted amino acid sequences of the DMS4 protein in the region affected by mutations (beginning at N121) in the *dms4-1* mutant and three suppressor mutants. Red letters indicate changed amino acids and black letters indicate amino acids that are identical to the WT sequence. (D) Detection of the DMS4 protein by Western blotting in WT T+S, *dms4-1*, and the four suppressor mutants. In the *dms4-1s4* suppressor mutant, the DMS4 protein is somewhat larger and much less abundant, which is consistent with the additional 48 nucleotides in the relatively rare cDNA containing a restored reading frame (Figure S2). Bottom shows the stained membrane as a loading control.

overlooked at that time, perhaps owing to the high density at which the M_1 plants were grown.

Nature of dms4-1 intragenic suppressor mutations

Three of the suppressor mutations, *dms4-1s1*, *dms4-1s2*, and *dms4-1s3*, created new splice site acceptors that were predicted to reestablish the wild-type *DMS4* open reading frame (Figure 3B). Sequencing of *dms4* cDNAs synthesized from mRNA isolated from each suppressor mutant confirmed that the new splice site acceptors were used, although there was some sequence variability among the cDNA clones due to alternative splicing (Figure S2). Despite minor variations in the amino acid sequence arising from use of the new splice site acceptors (Figure 3C), the DMS4 proteins translated from the resulting cDNAs could be detected on Western blots using a DMS4 polyclonal antibody (Figure 3D) and were fully functional in RdDM/GFP silencing and development, as demonstrated by the phenotypic analyses described above. One suppressor mutation, *dms4-1s4*, did not create a new canonical splice site acceptor site (Figure 3A) although a rare cDNA that has a nearly wild-type open reading frame was detected in this strain

Table 1 Mutation frequencies of suppressors

Allele	M_1 generation	M_2 generation	Total	Mutation frequency	Poisson
First mutagenesis (10,599)					
dms4-1s1	10	17	27	2.5×10^{-3}	3.1*E*-43
dms4-1s2	4	6	10	9.4×10^{-4}	1.0*E*-12
dms4-1s3	4	4	8	7.5×10^{-4}	1.0*E*-09
dms4-1s4	1	5	6	5.7×10^{-4}	6.7*E*-07
Total	19	32	51		
Second mutagenesis (3575)					
dms4-1s1	7	1	8	2.2×10^{-3}	2.1*E*-13
dms4-1s2	1	0	1	2.8×10^{-4}	9.0*E*-02
dms4-1s3	4	0	4	1.1×10^{-3}	3.7*E*-06
dms4-1s4	2	2	4	1.1×10^{-3}	3.7*E*-06
Total	14	3	17		

Two rounds of EMS mutagenesis were carried out on *dms4-1* seeds of the BC_1F_3 and BC_2F_4 generations, respectively. The number of M_1 plants obtained following each mutagenesis treatment is shown in parentheses. The number of times each suppressor mutation was identified in the M_1 and M_2 generations is indicated together with the mutation frequencies (calculated as the total number of plants having a mutation divided by the size of the M_1 population). The average mutation frequencies in the first and second round are 1.2×10^{-3} and 1.17×10^{-3}, respectively. Poisson probability indicates the probability of recurrent suppressor mutations. λ is calculated as (1189.8 (average number of induced mutations) $\times$ 10,599 or 3575 (mutagenized population)/42,859,753 (GCs in *Arabidopsis* genome).

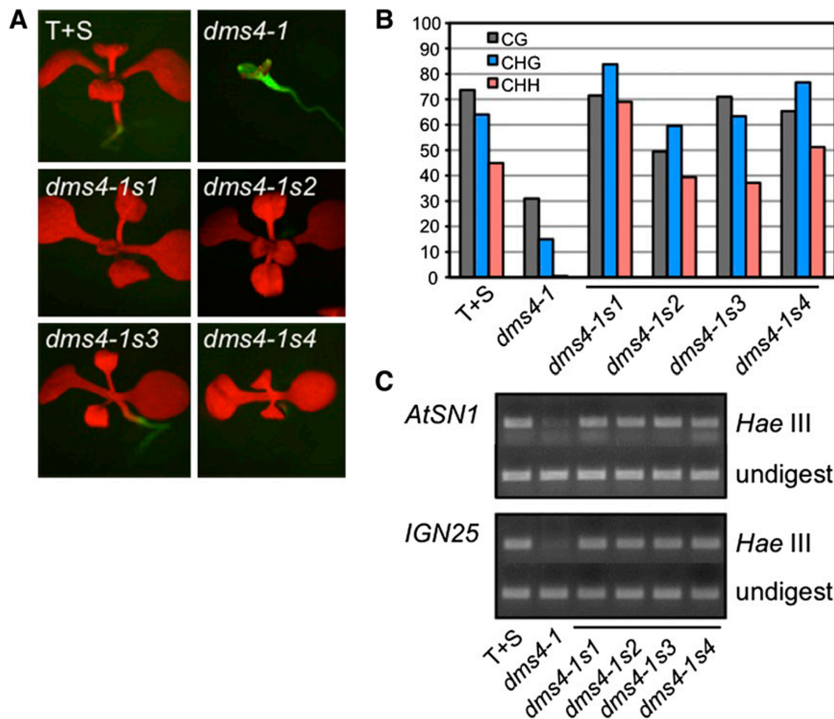


Figure 4 GFP signal and DNA methylation. (A) *GFP* is silenced in WT T+S seedlings, reactivated in the *dms4-1* mutant (Kanno *et al.* 2010), and resilenced in suppressor mutants *dms4-1s1* to *dms4-1s4*. The suppressor mutant seedlings also show a wild-type phenotype compared to the *dms4-1* seedling, which is delayed in germination and development (seedlings were photographed on the same day after sowing on MS medium; the *dms4-1* mutant is delayed in development relative to the other genotypes and shows *GFP* expression in the hypocotyl at this stage (Kanno *et al.* 2010). (B) Percentages of DNA methylation in all sequence contexts (CG, CHG, CHH, where H is A, T, or C) in the target enhancer region in WT T+S, *dms4-1*, and suppressor mutants as analyzed by bisulfite sequencing. Following loss of methylation in the *dms4-1* mutant, methylation is essentially restored to wild-type levels in the suppressor mutants. (C) Chop–PCR analysis to detect methylation in endogenous targets of RdDM. Genomic DNAs digested by *Hae*III were used as templates. Presence or absence of the amplification product in *Hae*III panels indicates the presence or absence of methylation, respectively.

(Figure S2). Accordingly, trace amounts of protein interacting with the DMS4 antibody were observed on the Western blot (Figure 3D). Thus, all phenotypes conditioned by the *dms4-1* mutation are corrected by the four intragenic suppressor mutations, which reestablish the correct reading frame of the *DMS4* gene.

High frequency of suppressor mutations

A total of 51 independent cases of the four *dms4-1* suppressor mutations were recovered from an initial population of ~10,600 M_1 plants (Table 1). The individual mutation frequencies (the number of times a mutation is observed divided by the size of the M_1 population) (Jander *et al.* 2003) ranged from 2.5×10^{-3} for *dms4-1s1* to 5.7×10^{-4} for *dms4-1s4*. The average mutation frequency was 1.2×10^{-3} (Table 1). These frequencies can be compared to those found in a previous case of saturation EMS mutagenesis in *Arabidopsis*, where the average frequency of mutations in the *CSR* (*CHLORSULFURON/IMIDAZOLINONE RESISTANT 1*) gene leading to herbicide resistance was determined (Jander *et al.* 2003). The *CSR* gene, which is 2700 bp in length and free of introns, encodes the catalytic subunit of acetolactate synthase (ALS). Four amino acid substitutions resulting from distinct point mutations in the *CSR* gene prevent binding of various ALS-inhibiting herbicides and hence confer herbicide resistance. The average mutation frequency observed in that study was 1.6×10^{-5} (Jander *et al.* 2003). The mutation frequencies we observed are thus ~35–150 times higher than reported previously for a case of standard EMS mutagenesis in *Arabidopsis*.

The high mutation frequencies of the *dms4-1* suppressor mutations were reproduced in a second trial of EMS mutagenesis using a more advanced generation of *dms4-1* seeds (BC_2F_4). In a population of 3575 M_1 plants and their M_2 progeny, we retrieved 17 plants containing one of the four *dms4-1* intragenic suppressor mutations identified previously, again giving an average mutation frequency of $\sim 1.2 \times 10^{-3}$ (Table 1).

Mutation frequency is not elevated genome-wide

The high frequencies of *dms4-1* intragenic suppressor mutations suggested the existence of a hypermutational process that affected the *dms4* gene. To test whether a similar elevated mutation frequency would be observed genome-wide, we carried out Illumina whole-genome sequencing on two independently derived strains from the *dms4-1s1* and *dms4-1s3* suppressor mutants, respectively, and one strain from the *dms4-1s4* suppressor mutant. For comparison, we sequenced the genome of the *dms4-1* mutant from the BC_1F_3 generation, which was used for the first EMS mutagenesis treatment in the suppressor screen (Figure 1A), as well as the wild-type transgenic line containing the *GFP* target locus and silencer locus (T+S) that was used in the original forward screen that identified the *dms4-1* mutant (Kanno *et al.* 2010).

Once a list of single nucleotide polymorphisms (SNPs) was established for each line, we subtracted SNPs present in the wild-type T+S transgenic line and in the *dms4-1* mutant. The remaining SNPs were considered to be ones that were induced by EMS mutagenesis during the suppressor screen. From this analysis, the total number of mutations induced by EMS treatment of *dms4-1* seeds ranged from 826 (*dms4-1s1*, strain 12-1-3) to 2140 (*dms4-1s3*, strain 3-2-3) (Table 2). These numbers are within the range reported previously

Table 2 Number of EMS-induced mutations in suppressor mutants

	Chr 1	Chr 2	Chr 3	Chr 4	Chr 5	Total
dms4-1s1 (6-3-5)	169	183	219	216	266	1053
dms4-1s1 (12-1-3)	192	123	182	170	159	826
dms4-1s3 (3-2-3)	448	331	442	362	557	2140
dms4-1s3 (3-4-2)	295	151	284	143	166	1039
dms4-1s4 (26-4a)	242	94	199	160	196	891
Average	269.2	176.4	265.2	210.2	268.8	1189.8

Whole-genome sequencing was performed on five suppressor mutants (strain number in parentheses) containing the indicated suppressor mutations (left). The number of G/C to A/T mutations on each chromosome as well as the total number of G/C to A/T changes genome-wide (ranging from 826 to 2140) is shown.

for EMS mutagenesis in *Arabidopsis* (Jander *et al.* 2003) and hence do not support the occurrence of genome-wide hypermutation in the suppressor mutants. Indeed if hypermutation were occurring throughout the genome at the same frequency observed for the *dms4-1* gene, then we should have detected ~53,000 mutations in each suppressor mutant [average mutation frequency of *dms4-1* suppressor mutations (1.2×10^{-3}; Table 1) times 4.4×10^{7}, which is the number of base pairs susceptible to EMS mutagenesis in *Arabidopsis* (Jander *et al.* 2003)].

The mutations were distributed throughout the genome and they affected different categories of sequences (*e.g.*, genes, transposons, pseudogenes) at approximately the same percentage at which they are represented in the *Arabidopsis* genome (Figure 5). The neighboring nucleotides of the G-to-A transitions induced by EMS treatment are consistent with previously reported results (Greene *et al.* 2003), with purines being favored in the −1 position (Table S4). There was virtually no overlap among the sets of genome-wide mutations detected in each suppressor mutant; that is, each mutation was induced independently and only once in the M_1 seed population. This contrasts to the suppressor mutations, which were induced independently multiple times (Table 1). We confirmed a subset of strain-specific SNPs by using cleaved amplified polymorphic sequence (CAPS) markers (Figure S3).

The dms4-1 allele is not a general target of enhanced mutagenesis

The four suppressor mutations were straightforward to detect because they led to reversion of *dms4-1* mutant phenotypes (that is, the suppressor mutants were GFP^- and had a normal developmental phenotype). To determine whether other G residues in the *dms4-1* allele were mutated frequently even in the absence of phenotypic reversion, we sequenced the *dms4-1* gene in 100 GFP^+ M_2 plants. No additional mutations were observed in the *dms4-1* allele in any of these plants, indicating that the *dms4-1* allele as a whole is not a preferential target of mutagenesis.

Mutations in the GFP reporter gene

We also determined the frequency of recessive, loss-of-function mutations occurring in the *GFP* reporter gene. These mutants could be screened out because they were GFP^- but had a *dms4-1* developmental phenotype. Sixteen independent GFP^- mutants resulting from mutations in the *GFP* coding sequence were obtained in the initial screen of M_2 progeny from ~10,600 M_1 plants resulting from the first EMS treatment. Although most mutations were observed only once, two were observed multiple times. The resulting mutation frequencies (1.9×10^{-4} and 2.8×10^{-4}, respectively) are somewhat elevated over previously reported average values (Jander *et al.* 2003) but do not reach the higher frequencies observed with the *dms4-1* allele (Figure S4).

Discussion

In a screen for genetic suppressors of the *dms4-1* mutation, which conditions defects in both RdDM/GFP silencing and plant development, we identified four dominant intragenic suppressor mutations. The suppressor mutations correct all of the *dms4-1* mutant phenotypes and hence confirm that both the developmental abnormalities and epigenetic deficiencies observed in the *dms4* mutant are due solely to the *dms4-1* mutation. The remarkable aspect of this suppressor

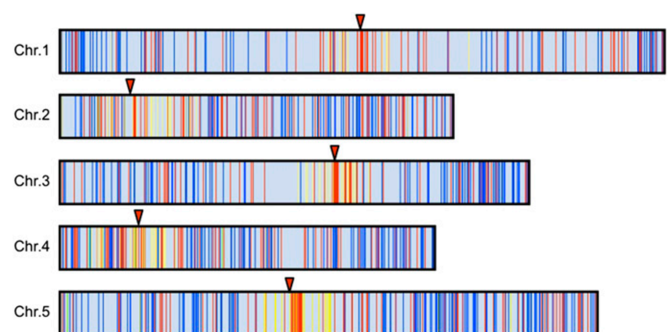


Figure 5 Distribution of EMS-induced mutations in suppressor mutants. The positions of unique SNPs on the five chromosomes of suppressor strain *dms4-1s1* (6-3-5) are indicated by vertical lines. Similar distributions of unique SNPs were observed in the other suppressor mutants subjected to whole-genome sequencing (not shown). SNPs are distributed among different types of sequences according to their approximate percentage of representation in the *A. thaliana* genome: genes (blue; 50%), intergenic (red; 40%), pseudogenes (green; 1%), and transposons (yellow; 8%). Red arrowheads indicate the position of centromeres. Some clustering of SNPs in transposons and intergenic regions is seen in centromeric regions.

screen, however, was the high frequency with which we recovered the suppressor mutations. The average frequency of the four *dms4-1* suppressor mutations was ~100 times higher than normally observed with standard EMS mutagenesis in *Arabidopsis*. As revealed by whole-genome sequencing, the mutation frequency was not elevated genome-wide in the suppressor mutants. Rather, four specific sites in the *dms4-1* allele appeared to be the targets of recurrent mutational events that were recovered at high frequency in our screen. Whole-genome sequencing confirmed that the four suppressor mutations were induced independently because each suppressor mutant contained, in addition to the specific suppressor mutation, hundreds of additional unique mutations. Moreover, the repeated and independent induction of the suppressor mutations was illustrated by their presence in multiple M_1 plants as well as M_2 progeny grown from separate pools of seeds.

Our results are reminiscent of those reported in a previous study on hypermutation in the *bal1* variant in *Arabidopsis*. The *bal1* variant arose during inbreeding of the epigenetic mutant *ddm1*, which is defective in a chromatin remodeler required for DNA methylation of repeated sequences. The *bal1* allele conditions a dwarf phenotype, which is due to constitutive overexpression of the *SNC1* gene that is part of a resistance gene cluster. Following treatment with EMS, the *bal1* variant was phenotypically unstable, with nearly one-third of M_1 plants exhibiting wild-type sectors (Yi and Richards 2008). An epigenetic source was initially suspected to account for the high frequency of phenotypic instability. However, subsequent work demonstrated a genetic mechanism involving duplication of the *SNC1* gene followed by an apparent hypermutational process that induced inactivating missense mutations in one copy, thus returning *SNC1* expression to a normal level. Mutagenesis appeared to be restricted to the *SNC1* duplication because a sequenced control region lacked mutations (Yi and Richards 2009).

To explain the high incidence of phenotypic suppression in *bal1* variants, the authors proposed two hypothetical mechanisms that are not mutually exclusive (Yi and Richards 2009). The stress-induced mutagenesis hypothesis invoked DNA damage induced by EMS acting as the stressful agent. The meristem selection hypothesis proposed that cells carrying a mutation in *SNC1* outcompete wild-type cells in the SAM. These hypothetical mechanisms can also be applied to our results by positing targeted mutagenesis of the *dms4-1* allele or selection of revertant cells containing a *dms4-1* suppressor mutation in the SAM.

Targeted mutagenesis implies that specific nucleotides are preferentially mutated at high frequencies. It is not clear why the four sites sustaining suppressor mutations in the *dms4-1* allele would be preferential targets of mutagenesis. The wild-type *DMS4* gene, which is present as a single copy in *Arabidopsis*, is not a preferential target of EMS-induced mutagenesis in wild-type plants. In the initial screen for *dms* mutants in a population of ~52,000 M_1 plants, we identified only two loss-of-function alleles in the *DMS4* gene (Kanno *et al.* 2010). By contrast, in the current *dms4-1* suppressor screen, the intragenic suppressor mutations were identified 51 and 17 times, respectively, in populations of only ~10,600 and 3575 M_1 plants. The nature of the *dms4-1* mutation, which is in a splice-site acceptor of the sixth intron, may be relevant. Three of four intragenic suppressor mutations create new splice site acceptors that override the original *dms4-1* mutation, leading to restoration of the wild-type open reading frame. Analysis of cDNAs in the *dms4-1* mutant (Kanno *et al.* 2010) and the suppressor mutants (this study) revealed the occurrence of alternative splicing. A speculative idea is that oscillations between different splicing site acceptors may preferentially fix mutations at specific splice sites in the *dms4-1* allele if they result in a mRNA encoding a functional DMS4 protein. Splicing-related parameters have been correlated with the selection of mutations in p53 in human cancers (Kouidou *et al.* 2009). The somewhat elevated frequency of mutation at two sites in the *GFP* reporter may appear to support targeted mutagenesis at some sites in the genome. Owing to its transgenic nature, however, it is difficult to judge the relevance of the results on the *GFP* reporter gene for mutation frequencies in endogenous genes.

A common feature of the *bal1* variant and the *dms4* mutant is that both are likely to harbor widespread epigenetic alterations in their genome. As mentioned above, *bal1* arose in the epigenetic mutant *ddm1*, which accrues epigenetic and genetic alterations during inbreeding. DMS4 can be considered an epigenetic factor because it directly or indirectly affects RdDM (Kanno *et al.* 2010). Because active demethylation of DNA is a base-excision repair process that can potentially be mutagenic (Zhu 2009), the epigenetic state of genes may make them more or less susceptible to EMS mutagenesis. However, we did not observe any differences in the DNA methylation state of the *dms4-1* allele compared to that of the wild-type *DMS4* gene (Figure S5).

Perhaps a more plausible explanation than targeted mutagenesis to account for the frequent recovery of the four suppressor mutations is that they restore DMS4 function and a normal plant phenotype. DMS4 is important for development and *dms4-1* suppressor mutations that reestablish a wild-type phenotype may provide a selective advantage over mutant cells in the SAM. Such intraorganismal selection is referred to as somatic, cell-lineage or diplontic selection (Clarke 2011). The young SAM in *Arabidopsis* consists of 50–70 cells (Medford 1992) arranged in several layers. The L1 layer forms the colorless epidermis, the L2 layer forms the subepidermis and germ cells, and the inner L3 region forms core tissues. Cells in L1 and L2 divide in an anticlinal (sideways) manner, which normally ensures that these layers are maintained separately from each other, whereas cell division in L3 occurs more randomly. Despite such compartmentalization, occasionally divisions can take place such that daughter cells invade a different

layer (Carpenter and Coen 1995; Clarke 2011). The stratified SAM of flowering plants would appear to reduce diplontic selection against deleterious mutations (Klekowski 2003) but there is little information on the positive selection of beneficial mutations in the SAM (Carpenter and Coen 1995; Szymkowiak and Sussex 1996; Klekowski 2003; McKey *et al.* 2010).

For diplontic selection to explain our results, a single revertant cell resulting from EMS mutagenesis would have to rapidly outcompete *dms4-1* mutant cells and spread to occupy a large portion of the SAM. This presumably takes place during early growth of the M_1 seedling because at the adult stage, M_1 suppressor mutants appear completely wild-type or contain obvious wild-type sectors. Moreover, the suppressor mutations are consistently inherited in M_2 progeny, indicating they were present in the pair of germ cell progenitors in the L2 layer of the SAM of M_1 plants (Yi and Richards 2008). The delay in germination and development seen in the *dms4-1* mutant at the seedling level (He *et al.* 2009; Kanno *et al.* 2010) can possibly be extrapolated to infer a growth advantage of revertant cells over *dms4-1* mutant cells in the SAM. In this context, it is interesting to note that nearly 130 cell cycle and cyclin-related genes are differentially regulated in the *dms4-1* mutant (Kanno *et al.* 2010), which might alter cell cycle control in a disadvantageous manner.

The precise mechanism(s) resulting in the frequent and recurrent recovery of *dms4-1* suppressor mutations remains to be clarified. However, our work and the previous study on the *bal1* variant suggest that *Arabidopsis* can frequently target or select for beneficial mutations during the lifetime of an individual plant and transmit these mutations to the next generation. The involvement of two different genes (*SNC1* and *dms4*) and two different types of mutations (missense mutations and splice site acceptor mutations) suggest that such phenomena are not restricted to a single system but may actually be quite common in plants. Although our experiments involve EMS mutagenesis of the *dms4-1* mutant, they nevertheless illuminate the possibility of positive selection acting on naturally induced genetic variation in the SAM and are consistent with considerable organizational flexibility of the SAM in *Arabidopsis* (Furner and Pumfrey 1992, 1993) when strong selective forces are at play. The *dms4* mutant provides a good system for further analysis of this apparent targeted mutation/selection process and its possible role in plant adaptation and evolution.

Acknowledgments

We thank the Next Generation Sequencing Core Facility at the Vienna Biocenter for assistance with whole-genome sequencing and Eric Richards, Scott Poethig, and Ueli Grossniklaus for helpful discussions. This work was supported by the Austrian Academy of Sciences.

Literature Cited

Carpenter, R., and E. S. Coen, 1995 Transposon induced chimeras show that floricaula, a meristem identity gene, acts non-autonomously between cell layers. Development 121: 19–26.

Clarke, E., 2011 Plant individuality and multilevel selection theory, pp. 227–250 in *The Major Transitions in Evolution Revisited*, edited by B. Calcott, and K. Sterelny. MIT Press, Cambridge.

Collins, S. R., K. M. Miller, N. L. Maas, L. Roguev, J. Fillingham *et al.*, 2007 Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature 446: 806–810.

Czeko, E., M. Seizl, C. Augsberger, T. Mielke, and P. Cramer, 2011 Iwr1 directs RNA polymerase II nuclear import. Mol. Cell 42: 261–266.

Earley, K. W., F. Pontvianne, A. T. Wierzbicki, T. Blevins, S. Tucker *et al.*, 2010 Mechanisms of HDA6-mediated rRNA gene silencing: suppression of intergenic Pol II transcription and differential effects on maintenance *vs.* siRNA-directed cytosine methylation. Genes Dev. 24: 1119–1132.

Esberg, A., Z. Moqtaderi, X. Fan, J. Lu, K. Struhl *et al.*, 2011 Iwr1 protein is important for preinitiation complex formation by all three nuclear RNA polymerases in *Saccharomyces cerevisiae*. PLoS ONE 6: e20829.

Eun, C., Z. J. Lorkovic, U. Naumann, Q. Long, E. R. Havecker *et al.*, 2011 AGO6 functions in RNA-mediated transcriptional gene silencing in shoot and root meristems in *Arabidopsis thaliana*. PLoS ONE 6: e25730.

Furner, I. J., and J. E. Pumfrey, 1992 Cell fate in the shoot apical meristem of *Arabidopsis thaliana*. Development 115: 755–764.

Furner, I. J., and J. E. Pumfrey, 1993 Cell fate in the inflorescence meristem and floral buttress of *Arabidopsis thaliana*. Plant J. 4: 917–931.

Greene, E. A., C. A. Codomo, N. E. Taylor, J. G. Henikoff, B. J. Till *et al.*, 2003 Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in Arabidopsis. Genetics 164: 731–740.

Haag, J. R., and C. S. Pikaard, 2011 Multisubunit RNA polymerases IV and V: purveyors of noncoding RNA for plant gene silencing. Nat. Rev. Mol. Cell Biol. 12: 483–492.

He, X. J., Y. F. Hsu, S. Zhu, H. L. Liu, O. Pontes *et al.*, 2009 A conserved transcriptional regulator is required for RNA-directed DNA methylation and plant development. Genes Dev. 23: 2717–2722.

Jander, G., S. R. Baerson, J. A. Hudak, K. A. Gonzalez, K. J. Gruys *et al.*, 2003 Ethylmethanesulfonate saturation mutagenesis in Arabidopsis to determine frequency of herbicide resistance. Plant Physiol. 131: 139–146.

Kanno, T., E. Bucher, L. Daxinger, B. Huettel, G. Böhmdorfer *et al.*, 2008 A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation. Nat. Genet. 40: 670–675.

Kanno, T., E. Bucher, L. Daxinger, B. Huettel, D. P. Kreil *et al.*, 2010 RNA-directed DNA methylation and plant development require an IWR1-type transcription factor. EMBO Rep. 11: 65–71.

Klekowski, E., 2003 Plant clonality, mutation, diplontic selection and mutational meltdown. Biol. J. Linn. Soc. Lond. 79: 61–67.

Kouidou, S., A. Malousi, and N. Maglaveras, 2009 Li-Fraumeni and Li-Fraumeni-like syndrome mutations in p53 are associated with exonic methylation and splicing regulatory elements. Mol. Carcinog. 48: 895–902.

Law, J. A., A. A. Vashisht, J. A. Wohlschlegel, and S. E. Jacobsen, 2011 SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling

factors, associate with RNA polymerase IV. PLoS Genet. 7: e1002195.
Matzke, M., T. Kanno, L. Daxinger, B. Huettel, and A. J. M. Matzke, 2009 RNA-mediated chromatin-based silencing in plants. Curr. Opin. Cell Biol. 21: 367–376.
McKey, D., M. Elias, B. Pujol, and A. Duputié, 2010 The evolutionary ecology of clonally propagated domesticated plants. New Phytol. 186: 318–332.
Medford, J. I., 1992 Vegetative apical meristems. Plant Cell 4: 1029–1039.
Reinders, J., C. Delucinge Vivier, G. Theiler, D. Chollet, P. Descombes *et al.*, 2008 Genome-wide, high-resolution DNA methylation profiling using bisulfite-mediated cytosine conversion. Genome Res. 18: 469–476.
Szymkowiak, E. J., and I. M. Sussex, 1996 What chimeras can tell us about plant development. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47: 351–376.
Yi, H., and E. J. Richards, 2008 Phenotypic instability of *Arabidopsis* alleles affecting a disease *Resistance* gene cluster. BMC Plant Biol. 8: 36.
Yi, H., and E. J. Richards, 2009 Gene duplication and hypermutation of the pathogen *Resistance* gene *SNC1* in the Arabidopsis *bal* variant. Genetics 183: 1227–1234.
Zhu, J. K., 2009 Active DNA demethylation mediated by DNA glycosylases. Annu. Rev. Genet. 43: 143–166.

Communicating editor: *T. Wu*

GENETICS

Supporting Information
http://www.genetics.org/content/suppl/2012/09/28/genetics.112.144634.DC1

Unusual Case of Apparent Hypermutation in *Arabidopsis thaliana*

Taku Sasaki, Ulf Naumann, Petar Forai, Antonius J. M. Matzke, and Marjori Matzke

Copyright © 2012 by the Genetics Society of America
DOI: 10.1534/genetics.112.144634

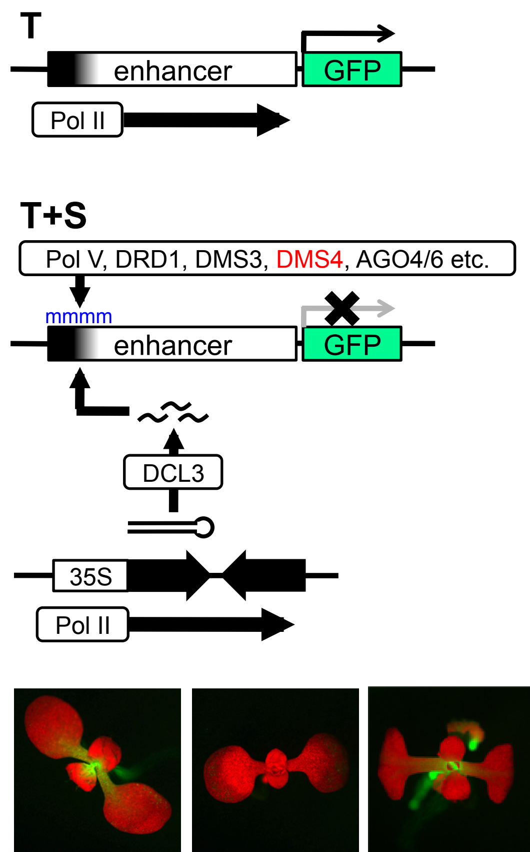


Figure S1 Transgene-based meristem silencing system. In the two component transgene silencing system, a target (T) locus contains a *GFP* reporter gene downstream of a minimal promoter and an upstream enhancer that drives *GFP* expression in shoot and root meristem regions (left photo, T). An unlinked silencer locus (S) contains an inverted DNA repeat (heavy black arrows) of distal enhancer sequences (dark shade) that is transcribed by Pol II from the 35S promoter. The resulting hairpin RNA is processed by DCL3 to produce 24-nt siRNAs that induce DNA methylation of distal enhancer sequences through Pol V pathway components, including DRD1, DMS3, DMS4 and AGO4/6. Methylation silences *GFP* expression (middle photo, T+S). GFP silencing is released in Pol V pathway mutants, including *dms4-1* (right photo).

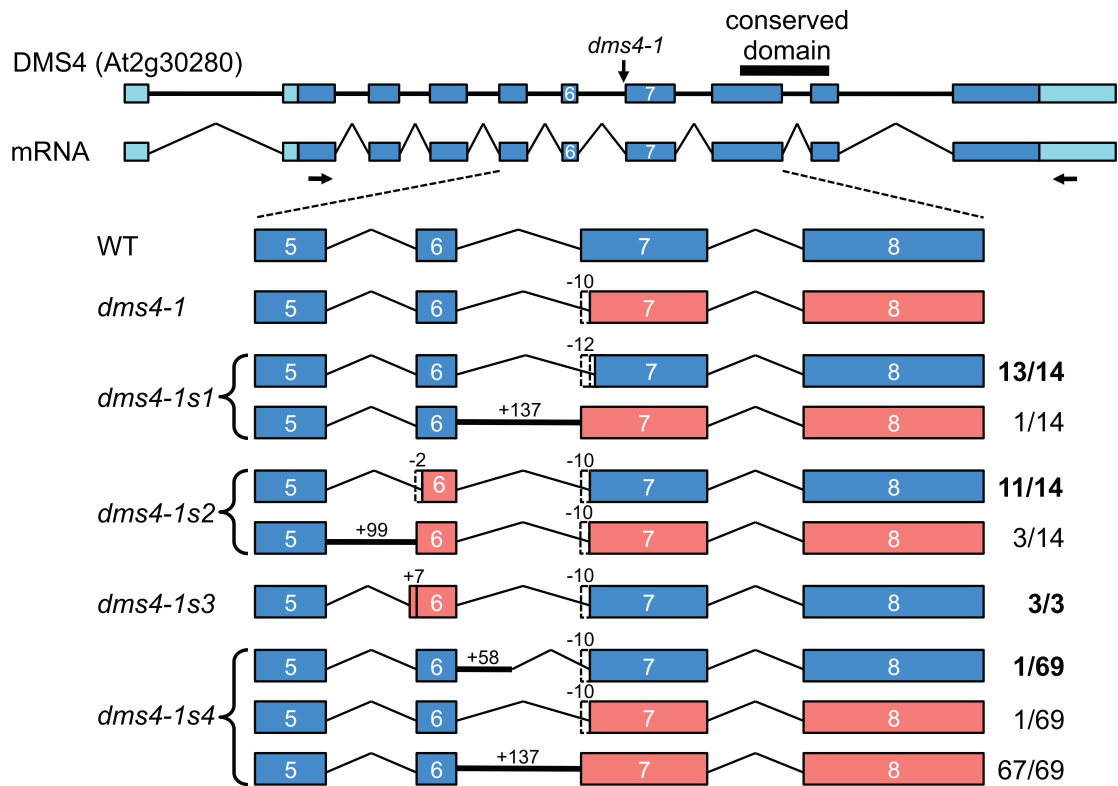


Figure S2 cDNA sequences of *DMS4* gene in suppressor mutants. The intron-exon structure of the *DMS4* gene is shown at the top. Below is the processed mRNA (spliced introns denoted by peaked lines). Light blue boxes denote UTRs, dark blue boxes coding sequences. The position of the *dms4-1* mutation (G to A at the splice site acceptor of the sixth intron) is indicated by the arrow. A conserved domain of the DMS4 protein is delineated by the black bar. Small arrows represent primers used for RT-PCR. Cloned *DMS4* cDNAs were sequenced from WT plants, the *dms4-1* mutant and the four suppressor mutants: *dms4-1s1* to *dms4-1s4*. Blue and red regions indicate those cDNAs with correct reading frames or incorrect reading frames, respectively (numbers of clones sequenced are shown as denominators to the right; the number with a correct reading frame is shown in bold). Horizontal bars denote unspliced intronic sequences. The gain (+) or loss (-) of nucleotides in mutant cDNAs are shown within unspliced intronic sequences or at the intron-exon junctions.

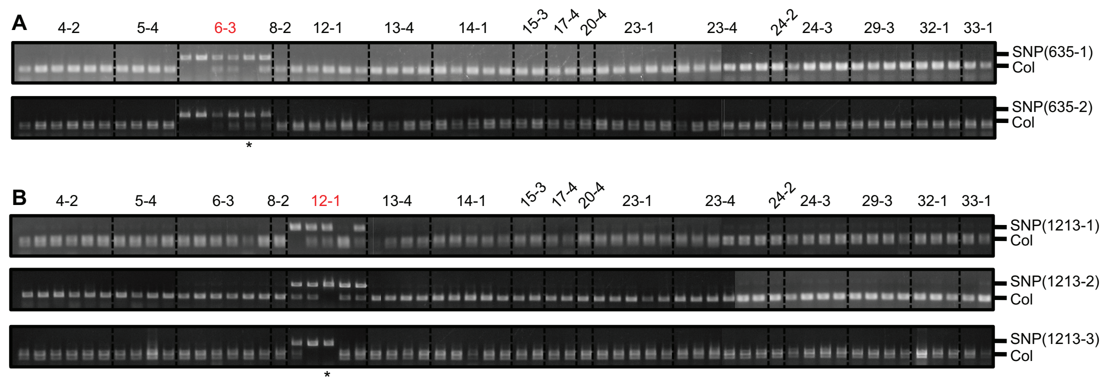


Figure S3 Genotyping of strain-specific SNPs in *dms4-1s1* by cleaved amplified polymorphic sequences (CAPS). To confirm that unique SNPs were indeed restricted to specific strains of suppressor mutants, CAPS markers were designed to detect several selected SNPs specific for *dms4-1s1* strain 6-3-5 (635-1 and 635-2) (A) and *dms4-1s1* strain 12-1-3 (1213-1 to 3) (B).

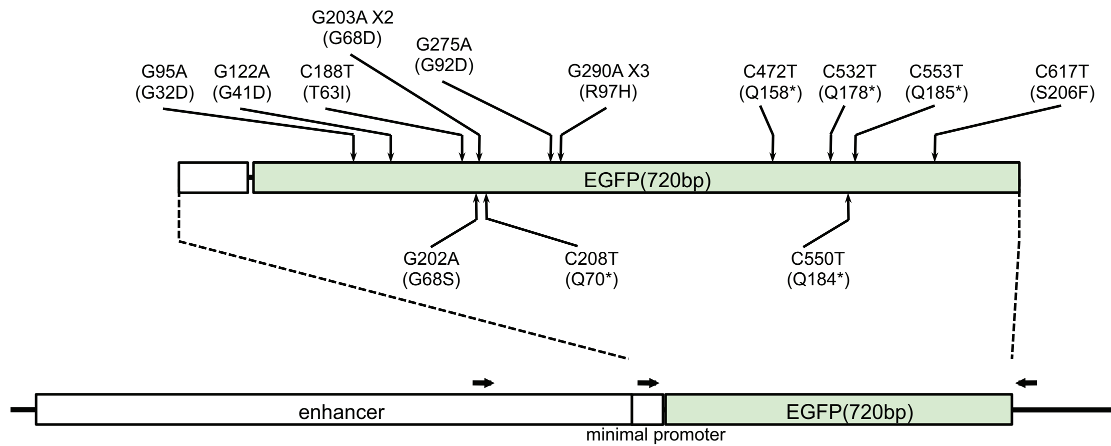


Figure S4 EMS-induced mutations in *GFP* reporter gene. The *GFP* reporter gene can be used to assess the mutation frequency of a selectively neutral gene. In the first EMS mutagenesis treatment, which involved ~ 10,600 M_1 plants, we identified 16 GFP-negative M_2 plants that had a *dms4-1*-like phenotype. In these cases, the GFP-negative phenotype was due to recessive loss-of-function mutations in the *GFP* gene itself. Although most of the mutations were observed only once (mutation frequency 1/10,600 or 9.4×10^{-5}), two were observed more than once: two times G203A (mutation frequency 2/10,600 or 1.9×10^{-4}) and three times G290A (mutation frequency 3/10,600 or 2.8×10^{-4}). These mutation frequencies are lower than those observed for the *dms4-1* suppressor mutations (Table 1).

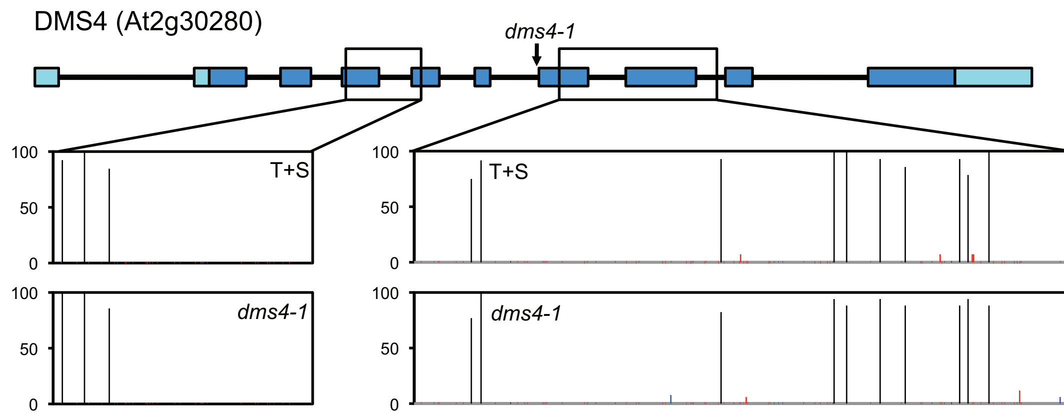


Figure S5 DNA methylation in the *DMS4* gene region around *dms4-1* mutation. Bisulfite sequencing was used to examine DNA methylation in the region around the *dms4-1* mutation. The *DMS4* gene contains CG methylation in the gene body (http://neomorph.salk.edu/epigenome/epigenome.html) and this methylation pattern appeared the same in the wild-type T+S line and the *dms4-1* mutant. Vertical bars indicate percent methylation (left) at Cs in CG dinucleotides in the boxed regions of the *DMS4* gene.

Table S1 Primers used in this study

primer name	sequence	Purpose
EPRV_Top2F	GCG GTG TYA TYT ATG TTA YTA GAT	Bisulfite for
EPRV_Top2R	CTT CTT RAT RTT CCA TAR CTT TCC	target
PHV_S-F2	GGA YYA TAG TGA TGY YAT ATT GTG	Bisulfite for
PHV_S-R	TAT CAT CAA CAA CTT TCC ACA CC	PHV
3028-3	GAA GCC TGT GAT TGT TAG AG	RT-PCR for
DMS4sqR2	CGG TAA TTC TCT TTA GTA TC	DMS4
AtSN1for	ACC AAC GTG CTG TTG GCC CAG TGG TAA ATC	chop-PCR for
AtSN1rev	AAA ATA AGT GGT GGT TGT ACA AGC	AtSN1
IGN25for	CTT CTT ATC GTG TTA CAT TGA GAA CTC TTT CC	chop-PCR for
IGN25rev	ATT CGT GTG GGC TTG GCC TCT T	IGN25
635-1f	GCG TCT ACC GTT TAG CGC TG	CAPS (*Nco* I)
634-1r	GCT TCT TCA GAC CCT CGA GG	for 635-1
635-2f	CTT TAG GGG TCT CAG TCT CC	CAPS (*Nco* I)
635-2r	GTA CAC CCG TAT GAT TCC TC	for 635-2
1213-1f	GAG CTT TAG GGA AGC TAA TC	CAPS (*Nco* I)
1213-1r	CCG CCA AGA AAC AGT GAC AG	for 1213-1
1213-2f	CTT CAG CCA GTC AGT TGC AC	CAPS (*Hind* III)
1213-2r	CAT GTT TCA CCA TCA CTA GC	for 1213-2
1213-3f	CGA TTT CGG AGT CGG AGT CG	CAPS (*Xho* I)
1213-3r	CCG TTG GAG GTA CAC TCT CC	for 1213-3

Table S2 Suppressor mutants screened from M_1 generation

strain name	mutation	M_2 segregation GFP$^-$	GFP$^+$	total
2-1a*	*dms4-1s2*	9	9	18
3-1a	*dms4-1s3*	30	7	37
3-2a	*dms4-1s1*	5	19	24
4-2a*	*dms4-1s1*	15	10	25
7-4a	*dms4-1s1*	16	5	21
8-1a*	*dms4-1s3*	15	7	22
9-2b	*dms4-1s2*	25	0	25
16-1a*	*dms4-1s1*	9	11	20
16-1b*	*dms4-1s1*	16	10	26
16-1c*	*dms4-1s1*	17	13	30
18-2a*	*dms4-1s1*	23	13	36
19-3a	*dms4-1s2*	11	9	20
24-1a	*dms4-1s1*	18	0	18
26-4a	*dms4-1s4*	30	6	36
30-2a	*dms4-1s1*	21	4	25
32-2a	*dms4-1s3*	5	20	25
32-3a	*dms4-1s2*	22	13	35
34-1a	*dms4-1s1*	15	10	25
35-3b	*dms4-1s3*	35	11	46

* Plants in which the *dms4* gene was sequenced in the M_1 generation; for the others, the *dms4* gene was sequenced in M_2 generation.

Nineteen suppressor mutants containing one of four dominant intragenic suppressor mutations (*dms4-1s1* to *dms4-1s4*) were identified in a population of 10,599 M_1 plants following the first EMS mutagenesis of *dms4-1* seeds (BC_1 F_3 generation). Selfed progeny of these plants (M_2 generation) segregated at least some GFP-negative progeny, indicating transmission of the suppressor mutations to the next generation and restoration of GFP silencing. Probably due to chimerism (Fig. 1A-C), segregation ratios of GFP-negative to GFP-positive were not always 3 to 1, as would normally be expected for progeny of an M_1 plant that is heterozygous for a dominant suppressor mutation. Two of the 19 M_1 plants, 9-2b and 24-1a, segregated 100% GFP-negative M_2 progeny. This result is difficult to explain unless the parental plants were homozygous for the suppressor mutation. This may have occurred through an early gene conversion event or if the same intragenic suppressor mutation was induced in both alleles of the *dms4-1* gene.

Table S3 Numbers of suppressor mutants screened from M_2 population

Batch No.	mutation	homozygous	heterozygous	screened No.
3-2	*dms4-1s3*	1	3	4
3-4	*dms4-1s3*	3	5	8
4-2	*dms4-1s1*	2	5	7
4-4	*dms4-1s4*	1	0	1
5-3	*dms4-1s2*	2	4	6
5-4	*dms4-1s1*	1	3	4
6-2	*dms4-1s3*	1	10	11
6-3	*dms4-1s1*	2	4	6
7-4	*dms4-1s1*	0	1	1
10-1	*dms4-1s2*	0	5	5
10-3	*dms4-1s4*	1	1	2
12-1	*dms4-1s1*	1	4	5
13-1	*dms4-1s4*	1	2	3
13-4	*dms4-1s1*	2	2	4
14-1	*dms4-1s1*	3	2	5
15-3	*dms4-1s1*	0	2	2
16-1	*dms4-1s2*	1	0	1
16-4	*dms4-1s4*	3	3	6
17-4	*dms4-1s1*	0	2	2
20-4	*dms4-1s1*	0	1	1
23-1	*dms4-1s1*	2	3	5
23-1	*dms4-1s3*	0	1	1
23-2	*dms4-1s2*	0	6	6
23-3	*dms4-1s2*	0	7	7
23-4	*dms4-1s1*	4	2	6
24-2	*dms4-1s1*	1	0	1
24-3	*dms4-1s1*	2	2	4
25-3	*dms4-1s2*	1	8	9
29-3	*dms4-1s1*	1	3	4
32-1	*dms4-1s1*	0	3	3
33-1	*dms4-1s1*	2	0	2
33-3	*dms4-1s4*	2	1	3
total				135

Approximately five-seven M_2 progeny (actual range 1-11) from each M_1 plant (population size 10,599) were germinated on solid MS medium and screened at the seedling stage for a GFP-negative phenotype, indicating restoration of *GFP* silencing. Thirty-two GFP-negative seedlings were identified. DNA sequence analysis of the *dms4-1* gene in these seedlings revealed that they were all homozygous or heterozygous for one of four dominant intragenic suppressor mutations (*dms4-1s1* to *dms4-1s4*). After transfer to soil, the GFP-negative seedlings all lacked features of the *dms4-1* developmental phenotype as adult plants.

Table S4 Frequencies of flanking sequences of mutated G residues (observed/expected)

Position	-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
A	0.89	0.96	0.87	0.88	1.09		1.08	0.88	0.94	0.93	0.87
T	0.88	0.92	0.91	0.94	0.52		0.67	0.82	0.83	0.94	0.88
C	1.25	1.10	1.07	1.42	1.07		1.31	1.02	0.94	1.16	1.25
G	1.16	1.12	1.33	0.92	1.64		1.15	1.51	1.47	1.08	1.21
X^2	145	39	170	228	908		351	366	308	48	180
P	2.5E-31	1.4E-08	1.2E-36	4.7E-49	1.3E-196		8.2E-76	5.2E-79	2.2E-66	2.0E-10	1.1E-38

EMS is an alkylating agent that targets G to produce O^6-ethylguanine, which is able to base pair with T but not C. During subsequent DNA repair, the original G/C pair can be replaced by A/T. In the table, zero is the position of the mutated G and the percentages of nucleotides downstream (-5 to -1) and upstream (+1 to +5) observed from whole genome sequencing of suppressor mutants are shown. The neighboring nucleotides of the G to A transitions induced by EMS treatment are consistent with previously reported results, with purines being favored in the -1 position (Greene et al., 2003).

Greene EA et al. (2003) Spectrum of chemically-induced mutations from a large-scale reverse-genetic screen in Arabidopsis. *Genetics* 164: 731-740.