SET DOMAIN GROUP2 is the major histone H3 lysine 4 trimethyltransferase in *Arabidopsis*

Lin Guo[a,1], Yanchun Yu[a,1], Julie A. Law[b], and Xiaoyu Zhang[a,2]

[a]Department of Plant Biology, University of Georgia, Athens, GA 30602; and [b]Department of Molecular Cell and Developmental Biology, University of California, Los Angeles, CA 90095

Edited* by Susan Wessler, University of California, Riverside, CA, and approved September 8, 2010 (received for review July 19, 2010)

Posttranslational modifications of histones play important roles in modulating chromatin structure and regulating gene expression. We have previously shown that more than two thirds of *Arabidopsis* genes contain histone H3 methylation at lysine 4 (H3K4me) and that trimethylation of H3K4 (H3K4me3) is preferentially located at actively transcribed genes. In addition, several *Arabidopsis* mutants with locus-specific loss of H3K4me have been found to display various developmental abnormalities. These findings suggest that H3K4me3 may play important roles in maintaining the normal expression of a large number of genes. However, the major enzyme(s) responsible for H3K4me3 has yet to be identified in plants, making it difficult to address questions regarding the mechanisms and functions of H3K4me3. Here we described the characterization of SET DOMAIN GROUP 2 (SDG2), a large *Arabidopsis* protein containing a histone lysine methyltransferase domain. We found that SDG2 homologs are highly conserved in plants and that the *Arabidopsis SDG2* gene is broadly expressed during development. In addition, the loss of SDG2 leads to severe and pleiotropic phenotypes, as well as the misregulation of a large number of genes. Consistent with our finding that SDG2 is a robust and specific H3K4 methyltransferase in vitro, the loss of SDG2 leads to a drastic decrease in H3K4me3 in vivo. Taken together, these results suggest that SDG2 is the major enzyme responsible for H3K4me3 in *Arabidopsis* and that SDG2-dependent H3K4m3 is critical for regulating gene expression and plant development.

chromatin | methylation

The nuclear DNA in eukaryotic cells is wrapped around histone octamers (two copies each of H2A, H2B, H3, and H4) to form nucleosomes, the basic units of chromatin. A myriad of molecular processes, such as transcription, replication, DNA repair, and recombination, take place in a chromatin environment and therefore can be profoundly affected by chromatin structure. A growing body of evidence from recent studies indicates that posttranslational covalent modifications of histones play important roles in modulating the structural properties of chromatin. Four types of histone modifications have been described in plants, namely methylation, acetylation, phosphorylation, and ubiquitination (1, 2). Adding to the complexity of histone modifications is the finding that at least two dozen residues in core histones can be modified (1).

Four lysine residues in *Arabidopsis* core histones can be methylated, specifically, histone H3 lysine4 (H3K4), H3K9, H3K27, and H3K36 (1). Results from recent genome-scale profiling studies suggest that most *Arabidopsis* genes contain methylated histones and that different types of methylation are preferentially associated with different transcription states (3–11). That is, actively transcribed genes usually contain H3K4me and H3K36me, whereas developmentally repressed genes are enriched for H3K27me3. These findings suggest that histone methylation may play broad and important roles in regulating transcription. Consistent with this notion, several *Arabidopsis* histone methyltransferase (HMTase) mutants have been found to exhibit a myriad of developmental abnormalities (12–26). Furthermore, transcriptome profiling studies have identified numerous genes that are misregulated in HMTase mutants (19, 27, 28).

Understanding of the activities of individual HMTases is crucial for elucidating the mechanisms and functions of histone methylation. More than 30 putative HMTase genes have been identified in the *Arabidopsis* genome (29–32). These genes are named *SET DOMAIN GROUP* (*SDG*) genes because they contain the SET lysine methyltransferase catalytic domain (named after the *Drosophila* histone methyltransferases *Su(var)3-9*, *Enhancer-of-zeste*, and *Trithorax*) (31). Previous phylogenetic analyses of plant SDGs have resolved five ancient classes (classes I–V) that have likely diverged before the diversification of plants, animals, and fungi, because all five classes also include animal and/or fungal proteins (31). In four of the five classes, the plant SDGs seem to share similar enzymatic activities with their animal and/or fungal homologs. Specifically, class I SDGs are involved in H3K27me2/3, class II in H3K36me2/3, class III in H3K4me2/3, and class V in H3K9me2 (13, 19, 33, 34). Class IV seems to be an exception: the *Arabidopsis* proteins SDG15 and SDG34 are responsible for H3K27me1 (24), but it is not clear whether the two yeast homologs (SET3 and SET4) possess HMTase activity. The HMTases responsible for H3K4me1, H3K9me1/3, and H3K36me1 have yet to be determined but are likely to be uncharacterized members in these five classes.

Our previous genome-wide study of H3K4me in *Arabidopsis* showed that more than two thirds of *Arabidopsis* genes contain at least one type of H3K4me and that actively transcribed genes are usually associated with H3K4me3 (9). These findings suggest that H3K4me3 may play important roles in maintaining the proper expression levels of numerous genes. However, how H3K4me3 HMTases are recruited to individual genes and the exact function(s) of H3K4me3 remain poorly understood. This largely reflects the fact that the major H3K4me3 HMTase has yet to be identified. Mutants in three of the seven class III HMTase genes (ATX1/SDG27, ATX2/SDG30, ATXR7/SDG25) display locus-specific defects in H3K4me; however, the global levels of H3K4me3 in these mutants remain similar to that of wild type (25, 26, 28, 35).

Here we describe the characterization of SDG2, a large and structurally unique protein belonging to class III. SDG2 shares no significant sequence homology with other *Arabidopsis* SDGs outside of the SET domain. However, we found that SDG2 homologs are widespread and highly conserved in plants and that *SDG2* is broadly expressed during *Arabidopsis* development. In addition, the loss of SDG2 leads to severe and pleiotropic phenotypes and the misregulation of numerous genes. Furthermore, SDG2 displays robust and specific H3K4me1, H3K4me2, and H3K4me3 HMTase activities in vitro. Finally, the loss of SDG2

Author contributions: L.G., Y.Y., and X.Z. designed research; L.G., Y.Y., and J.A.L. performed research; L.G., Y.Y., and X.Z. analyzed data; and X.Z. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

Data deposition: Raw and processed microarray data have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE23208).

[1]L.G. and Y.Y. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: xiaoyu@plantbio.uga.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1010478107/-/DCSupplemental.

GENETICS

is associated with a severe reduction in H3K4me3 in vivo. Taken together, these results indicate that SDG2 is likely the major enzyme responsible for H3K4me3 in *Arabidopsis*.

Results and Discussion

SDG2 Homologs Are Highly Conserved in Plants. Sequence analysis of an *SDG2* cDNA cloned from *Arabidopsis* indicated that the *SDG2* ORF is 7,245 bp in length, encoding a 2,414-aa protein (≈276 kDa) with the SET domain located between amino acids 1849 and 1989 (Fig. 1). The remaining region in SDG2 does not contain any recognizable domains or share any significant homology with other *Arabidopsis* proteins. Blastp searches using the SDG2 protein sequence as query against the National Center for Biotechnology Information (NCBI) nr protein database identified six potentially full-length proteins with extensive sequence similarity to SDG2 from a wide range of plant species, including monocots, dicots, moss, and the single-celled alga *Chlamydomonas reinhardtii* (referred to as SDG2 homologs; see below). Phylogenetic analysis showed that SDG2 and its homologs are more closely related to each other than to other proteins from the respective species (Fig. 1*A*). It is therefore likely that SDG2 and its homologs represent a distinct lineage of putative HMTases that have diverged from other HMTases during early plant evolution. Importantly, most SDG2 homologs identified here are computationally translated from mRNA sequences, indicating that they are encoded by transcribed genes in the respective species. Consistent with notion, further tBlastn searches against the est database recovered 51 additional hits (E value < 1e-100) from a wide range of plant species.

SDG2 and the six full-length homologs are similar in size (e.g., 2,300 aa in grapevine, 2,240 aa in rice, 2,852 aa in *Physcomitrella patens*) as well as structure with regard to the relative location of the SET domain. Sequence comparison by multiple alignment revealed a high level of conservation within the SET domain, as well as in the surrounding regions among SDG2 homologs (Fig. 1*B*). In addition, a number of highly conserved blocks of amino acids can be found throughout the length of these proteins, including the extreme N terminus (Fig. 1*B*). None of the conserved regions outside of the SET domain shows any significant homology with other SDG proteins in *Arabidopsis* or other proteins in the database. These results suggest that SDG2 might be functionally distinct from other characterized *Arabidopsis* HMTases.

SDG2 Is Broadly Expressed and Critically Important for Multiple Developmental Processes in *Arabidopsis*. *SDG2* transcripts can be readily detected in all major *Arabidopsis* organs by RT-PCR, including roots, stems, leaves, and inflorescence (Fig. S1). In addition, on the basis of a previously published transcriptome profiling dataset in 79 *Arabidopsis* tissue types, *SDG2* seems to be expressed relatively evenly in all tissues tested, at levels that are comparable to several other characterized HMTase genes (e.g., *SDG8*, *SDG15*, and *SDG33*) (36). To further characterize the expression pattern of SDG2, we generated transgenic plants expressing the β-glucuronidase (GUS) gene under the control of the *SDG2* promoter (the ≈1.3-kb genomic regions between *SDG2* and the upstream gene). Consistent with the RT-PCR results and previously published microarray data, multiple P_{SDG2}:*GUS* transgenic lines consistently exhibited a broad GUS staining pattern in multiple tissues throughout development (Fig. 2*A*).

The wide distribution and high level of sequence homology of SDG2 homologs in plants as well as the broad expression pattern of *SDG2* in *Arabidopsis* suggest that SDG2 may play important roles in regulating cellular or developmental processes. To test this, we obtained two T-DNA insertion mutant lines, *sdg2-1* (SALK_021008) and *sdg2-2* (CS852810), from the *Arabidopsis* Biological Research Center. Both T-DNA insertions are located upstream of the SET domain (Fig. 1*B*), and neither allele is capable of producing full-length transcripts as assayed by RT-PCR (Fig. S1). Heterozygous *sdg2-1*/+ and *sdg2-2*/+ plants are phenotypically indistinguishable from wild type, and homozygous *sdg2-1* and *sdg2-2* plants (referred to as *sdg2* below) are viable. However, both homozygous lines exhibit the same set of strong, pleiotropic phenotypes. The morphological defects in *sdg2* first become apparent at 6–8 d after germination, when *sdg2* seedlings are smaller with curly leaves and significantly shorter roots (Fig. 2*B*). *sdg2* plants remain dwarfed, with smaller rosettes throughout vegetative growth, and flower significantly earlier in all photoperiods tested (i.e., short day, long day, and constitutive light) (Fig. 2 *C* and *D*). This early transition from vegetative to reproductive growth is accompanied by the down-regulation of flowering repressor gene *FLC* (see below). Moreover, although the correct numbers of floral organs (sepal, petal, stamen, and carpel) are formed in *sdg2* flowers, *sdg2* mutants are completely sterile. A detailed examination revealed severe defects in both male and female reproductive organs in *sdg2*. Specifically, *sdg2* pollen grains are collapsed and unable to rehydrate or germinate when transferred to the stigma of wild-type plants. Nonetheless, *sdg2* pollen grains do seem to contain one vegetative and two generative nuclei (Fig. 2 *E* and *F* and Fig. S2). In addition, an embryo sac that normally contains seven cells is not formed in *sdg2*. Instead, a solid mass of small cells occupies the embryo sac (Fig. 2*G* and Fig. S2). Consistent with this phenotypic defect, the expression of egg- and endosperm-specific genes (e.g., *EC1* and *MEA*) is reduced to undetectable levels in *sdg2* (Fig. 2*H*) (37, 38). Finally, *sdg2* is defective in maintaining the indeterminacy of the inflorescence meristem. Instead of indeterminate growth, *sdg2* inflorescences eventually develop into a terminal flower with severely transformed organs (Fig. 2*I*).

The fact that the developmental abnormalities seen in *sdg2* are strictly recessive and shared by two independent T-DNA insertion mutants suggests that these phenotypes are caused by the loss of SDG2 function. To further test this, we performed a complementation test using the ≈13.5-kb genomic region containing the *SDG2* gene (including the ≈1.3-kb promoter, the ≈1-kb 3′ flanking region, and a 7x Myc epitope tag translationally fused to the N terminus of SDG2). Because *sdg2* plants were sterile, we transformed this construct into heterozygous *sdg2-1*/+ plants and identified T1 plants that were homozygous for *sdg2-1* and also

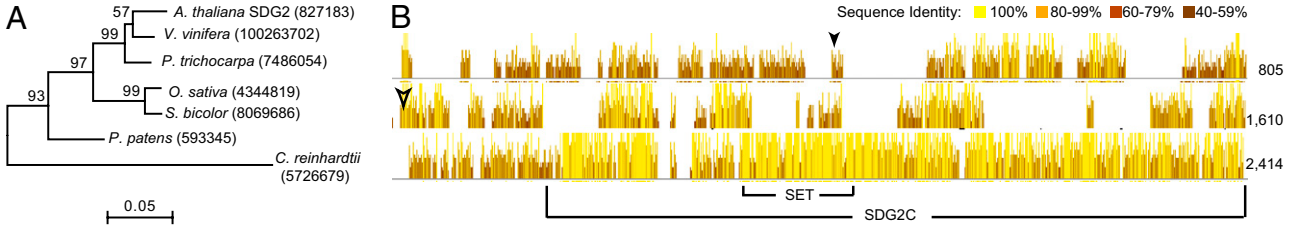

Fig. 1. SDG2 homologs in plants. (*A*) An unrooted neighbor-joining tree showing the phylogenetic relationship of SDG2 and full-length SDG2 homologs in plants. Numbers indicate bootstrap values. The GeneIDs are indicated next to the species names. Scale bar indicates sequence divergence. (*B*) A CLUSTALX histogram from a multiple alignment of the sequences in *A*. Numbers to the right indicate sequence positions (relative to SDG2). Vertical bars represent the sequence identity at each position (height and colors representing levels of identify). Open and filled arrowheads indicate T-DNA insertion positions in *sdg2-1* and *sdg2-2*, respectively. The positions of the SET domain and the SDG2C fragment are indicated below.

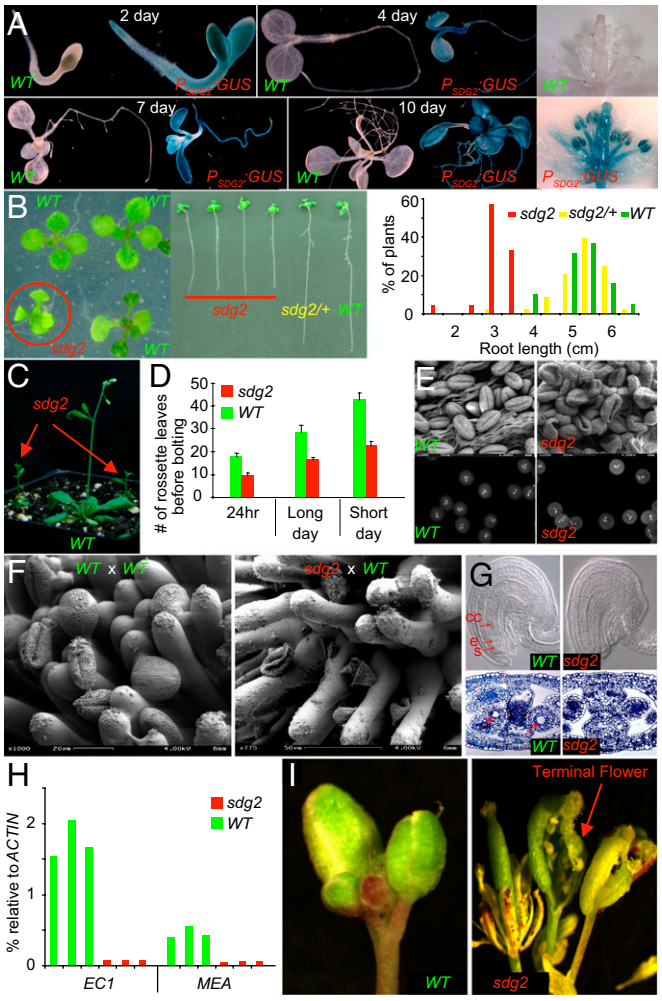

Fig. 2. Expression pattern of *SDG2* and pleiotropic phenotypes of *sdg2*. (*A*) GUS staining results of young seedlings and inflorescences of wild-type and P_{SDG2}:*GUS* transgenic plants. (*B*) *sdg2* seedlings are small, with curly leaves and reduced root length. (*C*) *sdg2* mutants remain dwarfed throughout development. (*D*) *sdg2* mutants flower early. (*E* and *F*) *sdg2* pollen grains are collapsed and unable to germinate when transferred to wild-type stigma, but the three haploid nuclei are present. (*G*) *Upper:* Longitudinal sections showing the central cell (cc), egg cell (e), and synergid cell (s) in wild-type but not in *sdg2* ovules. *Lower:* Cross-sections showing the embryo sac (arrows) in wild-type but not *sdg2* ovules. (*H*) Genes preferentially expressed in egg cells (e.g., *EC1*) and endosperm (e.g., *MEA*) are drastically down-regulated in *sdg2* (results from three replicates are shown). (*I*) *sdg2* inflorescences develop into terminal flowers with deformed organs.

contained the transgene (*sdg2-1;*P_{SDG2}*:myc-SDG2*) by PCR (Fig. S3*A*). Myc-SDG2 fusion proteins with the apparent molecular mass of ≈280 kDa can be readily detected in *sdg2-1;*P_{SDG2}*:myc-SDG2* plants by Western blots using an antibody against the Myc tag (Fig. S3*B*). Importantly, multiple independent transgenic plants were found to be phenotypically indistinguishable from wild type, indicating that the developmental abnormalities were fully rescued by the transgene. Taken together, these results indicate that SDG2 plays critical roles in regulating multiple developmental processes in *Arabidopsis*.

SDG2 Is Required for the Normal Expression of a Large Number of Genes. The pleiotropic defects of *sdg2* mutants suggest that SDG2 may be required for the normal expression of a large number of genes throughout development. To identify genes misregulated in *sdg2*, we performed a transcriptome analysis of *sdg2-1* and wild-type plants using the Affymetrix ATH1 microarrays (covering ≈22,600 *Arabidopsis* genes). *sdg2-1* and wild-type plants were grown side by side and analyzed 12 d after germination, when *sdg2-1* plants could be readily distinguished from wild type and heterozygous siblings.

We found that a large number of genes were misregulated in *sdg2-1*: 271 genes were up-regulated, and 321 genes were down-regulated by greater than fourfold (1,170 up-regulated and 1,265 down-regulated by greater than twofold). Real-time RT-PCR validations at 17 genes, using independently prepared samples, yielded results that were highly consistent with microarray results (Fig. S4). In some cases the gene expression alterations can already account for some of the phenotypic changes seen in *sdg2-1*. For example, consistent with the early flowering defect, the flowering repressor gene *FLC* was found to be significantly down-regulated in *sdg2-1* (At5g10140 in Fig. S4). Interestingly, whereas genes up-regulated in *sdg2-1* are involved in a wide range of biological processes, transcription factors and DNA-binding proteins are overrepresented in down-regulated genes (P = 3.9e-06 and 3.0e-04, respectively), suggesting that SDG2 might be required to maintain the proper expression level of many transcriptional regulators. Finally, although more than 10% of all genes are misregulated in *sdg2-1* at the seedling stage, it is likely that many additional genes may require SDG2 for their proper expression in other tissues (e.g., *EC1* and *MEA*; Fig. 2*H*). Taken together, these results suggest that SDG2 plays broad and important roles in regulating gene expression throughout development.

The substrate specificities and mutant transcriptional profiles have been previously described for two other *Arabidopsis* histone methyltransferases: ATX1/SDG27 (H3K4me) and the ASHH2/SDG8 (H3K36me) (19, 27). If SDG2 is functionally similar to ATX1 or ASHH2, one might expect similar transcriptional changes in these mutants. However, we found that the genes reported to be significantly up- or down-regulated in *atx1* and *ashh2* were not preferentially affected in the same manner in *sdg2-1* (Fig. S5). Although experimental variations such as genetic background or growth conditions may account for some of the observed differences in these mutants, the gross lack of correlation and the fact that *sdg2*, *atx1*, and *ashh2* mutants exhibit distinct sets of phenotypes suggest that SDG2, ATX1, and ASHH2 may play distinct roles in transcriptional regulation.

SDG2 Is a Histone H3 Lysine4 Methyltransferase. To determine whether SDG2 is indeed an active HMTase, we expressed and purified the C-terminal region of SDG2 including the SET domain and its flanking regions (amino acids 1650–2414; "SDG2C" in Fig. 1*B*). SDG2C was used here for two reasons. First, attempts to clone the full-length *SDG2* cDNA amplified by RT-PCR into *Escherichia coli* were unsuccessful, which might be because SDG2 was toxic in *E. coli*. Second, on the basis of the multiple alignment shown in Fig. 1*B*, SDG2C seems to represent a large and highly conserved domain, which is separated from other conserved regions by stretches of amino acids with low levels of sequence conservation. HMTase activity assays were performed using SDG2C as the enzyme source, calf thymus core histones or recombinant *Arabidopsis* histone H3 as the substrate, and S-[methyl-^{3}H]-adenosyl-L-methionine (SAM) as the methyl donor. As shown in Fig. 3*A*, SDG2C can readily methylate both calf thymus histone H3 and recombinant *Arabidopsis* histone H3. These results indicate that SDG2 is indeed an active H3 HMTase in vitro and that the HMTase activity of SDG2C does not require any preexisting histone modifications or other *Arabidopsis* proteins.

Arabidopsis histone H3 can be methylated at four lysines: H3K4, H3K9, H3K27, and H3K36 (1). To further characterize the substrate specificity of SDG2, we expressed and purified the N-terminal 57 aa of histone H3 as well as four mutant versions, each with one of the above four lysines mutated to an arginine (referred to as H3, H3K4R, H3K9R, H3K27R, and H3K36R, respectively, in Fig. 3*B*). Control experiments using the SET domain of the H3K9 HMTase KRYPTONITE (KYP) showed that KYP could methylate H3K4R, H3K27R, and H3K36R peptides but not the H3K9R peptide (Fig. 3*B*) (33). These mu-

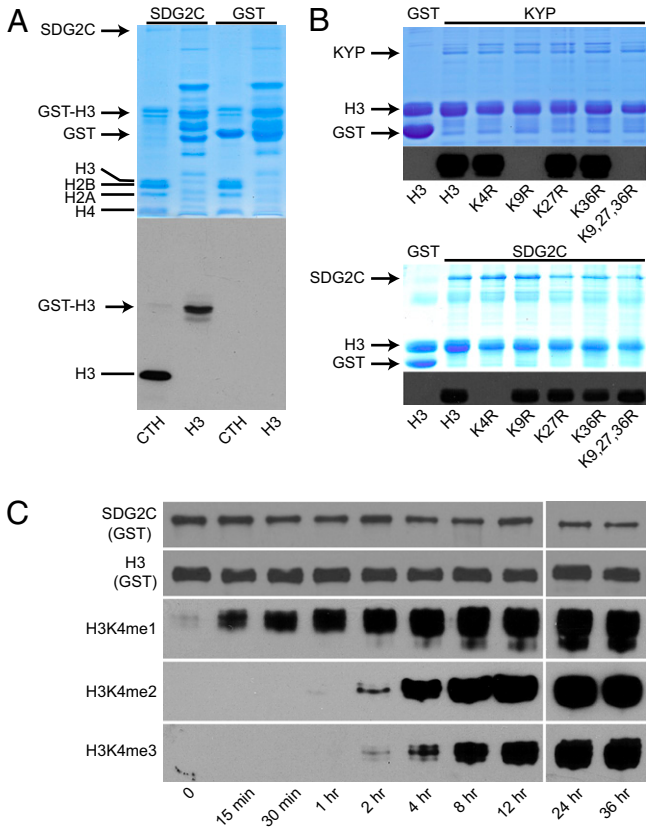

Fig. 3. In vitro HMTase activities of SDG2. (*A*) HMTase assay results using calf thymus histones (CTH) and recombinant *Arabidopsis* H3 (H3) as substrates. The proteins tested as enzyme sources are indicated on top, and the substrates are indicated below each lane. *Upper:* Coomassie blue-stained gel picture. *Lower:* Corresponding autoradiograph. (*B*) SDG2C specifically methylates H3K4. Coomassie blue-stained gel pictures (*Upper*) and the corresponding autoradiograph (*Lower*) are shown for assays using KYP and SDG2C as enzyme sources. Substrates (wild-type *Arabidopsis* $H3_{1-57}$ and K-to-R mutants) are indicated below each lane. (*C*) Western blot analysis showing the presence of H3K4me1, H3K4me2, and H3K4me3 in HMTase assay products. The durations of each HMTase assay are indicated below each lane, and the antibodies used are indicated on the left.

tant histones were then used as substrates for SDG2C. We found that H3K9R, H3K27R, and H3K36R peptides can be methylated by SDG2C at the same level as wild-type H3 (Fig. 3*B*). In addition, the simultaneous mutations of H3K9, H3K27, and H3K36 to arginines in the H3K9,27,36R triple-mutant peptide blocked methylation by KYP but not by SDG2C (Fig. 3*B*). In contrast, a single H3K4R mutation completely blocked methylation by SDG2C (Fig. 3*B*). These results indicate that SDG2C specifically methylates H3K4 in vitro.

To further characterize the activity of SDG2 with regard to the number of methyl groups added, we carried out HMTase assays using unlabeled SAM and allowed the reactions to proceed for various time periods from 15 min to 36 h. Western blot analyses were then performed using antibodies against H3K4me1, H3K4me2, and H3K4me3 to detect the presence of these modifications in the products. We found that SDG2C was capable of catalyzing all three types of H3K4 methylation: H3K4me1 first became detectable within 15 min and continued to accumulate throughout the 24–36-h span, whereas robust signals of H3K4me2 and H3K4me3 were detected after 2–4 h (Fig. 3*C*). Although the interpretation of these results is complicated by the semiquantitative nature of Western blot analysis as well as the in vitro conditions for the methylation reactions, it is interesting to consider that H3K4me2 and H3K4me3 catalyzed by SDG2 may occur at much slower rates than H3K4me1.

Loss of SDG2 Leads to a Severe and Genome-Wide Decrease in H3K4me3. The finding that SDG2C can catalyze all three types of H3K4 methylation is consistent with the fact that the SET domain of SDG2 contains a phenylalanine residue at an important position known as the "Phe/Tyr switch," a hallmark of HMTases capable of catalyzing all three types of lysine methylation (39). Importantly, of the seven *Arabidopsis* class III SDG proteins computationally identified as putative H3K4 HMTases, only SDG2 contains a Phe residue at the Phe/Tyr switch; the remaining six SDGs contain a Tyr residue at this position, a characteristic of HMTases that preferentially catalyze mono- and dimethylation.

To infer the in vivo specificity of SDG2, we assessed the levels of H3K4me1, H3K4me2, and H3K4me3 by Western blotting using proteins extracted from wild-type and *sdg2-1* plants. This analysis revealed a severe decrease in the cellular level of H3K4me3 in *sdg2-1* (but not in *sdg2-1;*P_{SDG2}*:myc-SDG2*) (Fig. 4*A*). Semiquantitative Western blot analysis showed that the cellular level of H3K4me3 in *sdg2-1* decreased to ≈34% of the wild-type level (Fig. S6). In contrast, H3K4me1 and H3K4me2 levels do not seem to be significantly affected in *sdg2* (Fig. 4*A*). Considering the fact that SDG2 can catalyze all three types of H3K4 methylation in vitro, the specific loss of H3K4me3 in *sdg2-1* suggests that the in vivo activity of SDG2 might be modulated by additional factors to favor H3K4me3. Alternatively, it is possible that the function of SDG2 in H3K4me1 and H3K4me2 is redundant with other class III SDGs. Future genetic studies involving multiple class III mutants should provide the necessary information to discern these two possibilities. It should also be noted that roughly one third of H3K4me3 remains in *sdg2*. This result is consistent with a previous report of a small decrease of H3K4me3 in the *atxr1* mutant (35). It is therefore likely that other class III HMTases can also function in the catalysis of H3K4me3, albeit with a significantly lower efficiency.

To further characterize the role of SDG2 in H3K4me3, we compared the chromatin levels of H3K4me3 between wild type and *sdg2-1* by ChIP and real-time PCR at 30 genes. These genes were selected to represent up-regulated, down-regulated, and unaffected genes, so that any changes in H3K4me3 caused by secondary effects (e.g., those associated with transcriptional changes) could be evaluated. The H3K4me3 level at each gene was determined as the ratio of H3K4me3 over H3 and normalized by the ratio of H3K4me3 over H3 at *Ta3*, a transcriptionally silent retrotransposon. In addition, an intergenic control region (between AT5G32950 and AT5G32975) was included in this analysis and found to contain virtually the same level of H3K4me3 as *Ta3* in both wild type and *sdg2-1* (Fig. 4*B*). The H3K4me3 levels at 5 of the 30 genes were relatively low in wild type but significantly higher (approximately two- to fivefold) than the intergenic region or *Ta3*. Loss of SDG2 did not lead to any significant decrease of H3K4me3 at these genes (Fig. 4*B*). In contrast, the H3K4me3 levels at the remaining 25 genes are much higher in wild type (≈10- to 100-fold of the intergenic region or *Ta3*) and become drastically reduced in *sdg2-1*, reaching levels that are comparable to the 5 genes described above (Fig. 4*B*). Importantly, the H3K4me3 defects at these genes are fully rescued by the P_{SDG2}*:myc-SDG2* transgene (Fig. 4*B*). These results are consistent with the drastic decrease (but not the elimination) of H3K4me3 detected by Western blot analysis (Fig. 4*A*). Taken together, these results indicate that SDG2 is responsible for the high levels of H3K4me3 at the vast majority of genes in *Arabidopsis*.

Interestingly, changes in H3K4me3 at individual genes do not seem to be strictly correlated with changes in the transcription levels of these genes. For example, the H3K4me3 levels at all 25 "high-H3K4me3" genes are drastically reduced in *sdg2-1*, regardless of whether a particular gene is up-regulated, down-regulated, or unaffected (Fig. 4*B*). It is therefore likely that the decrease of H3K4me3 at these genes results from the loss of

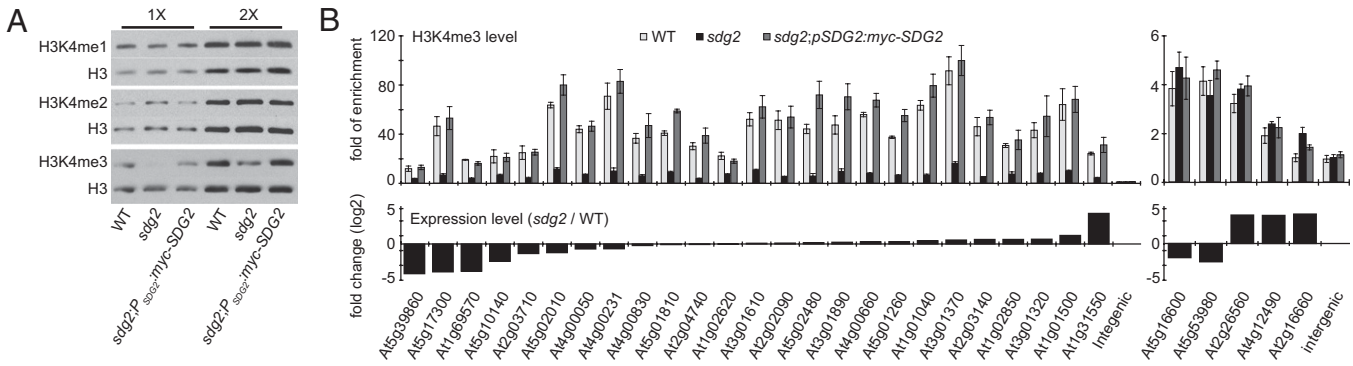

Fig. 4. H3K4me3 level is severely reduced in *sdg2*. (*A*) Western blot analysis of the cellular level of H3K4me1, H3K4me2, and H3K4me3 in wild-type, *sdg2-1*, and *sdg2-1;*P_{SDG2}*:myc-SDG2*. For each modification, the same membrane was stripped and blotted with an antibody against H3. Results are shown for two loading amounts that differ by twofold. (*B*) Level of H3K4me3 in wild-type, *sdg2-1*, and *sdg2-1;*P_{SDG2}*:myc-SDG2* plants from three replicates. *Left:* 25 genes with high H3K4me3 levels in wild type. *Right:* Five genes with low H3K4me3 levels in wild type. *y*-axis: fold of H3K4me3 enrichment over *Ta3*. Note that the *y*-axis scales are different between *Left* and *Right*. *Lower Left* and *Lower Right:* Transcriptional changes in *sdg2* for each gene. *y*-axis: fold change in *sdg2* relative to wild type ($\log_2$ scale).

SDG2, rather than from secondary effects associated with changes in transcription. The variable effects of the loss of H3K4me3 also indicate that the role of H3K4me3 in transcriptional regulation may be gene-specific and affected by additional factors.

Conclusions

In summary, two main lines of evidence presented here suggest that SDG2 is the major H3K4me3 HMTase in *Arabidopsis*. First, SDG2C specifically methylates H3K4 and is capable of catalyzing all three types of methylation in vitro. Second, the loss of SDG2 activity in *sdg2* leads to a severe decrease of H3K4me3 at numerous loci in vivo. These results are consistent with the fact that SDG2 is the only H3K4 HMTase in *Arabidopsis* that contains a Phe residue at the "Phe/Tyr switch," a characteristic of histone lysine trimethyltransferases (39). Importantly, the presence of a Phe residue at this position is conserved in all SDG2 homologs identified here, indicating that SDG2 homologs may also function in H3K4me3 in other plant species.

The identification of SDG2 as the major H3K4me3 HMTase in *Arabidopsis* should facilitate future studies directed to understand the mechanisms by which H3K4me3 is established at individual genes and to answer the question of how H3K4me3 may affect chromatin structure and transcription. Meanwhile, several findings described here have already broadened our understanding of H3K4me3 in plants. First, our previous genome-wide analysis provided circumstantial evidence that H3K4me may play broad and important roles in maintaining active transcription (9). The pleiotropic phenotypes and large-scale transcriptional changes associated with the loss of H3K4me3 in *sdg2* provide direct evidence that H3K4me3 is critically important in regulating gene expression in plants. Second, although the H3K4me3 levels at all 25 high-H3K4me3 genes decreased substantially, only a subset of these genes displayed significant changes in transcription, and both up- and down-regulation were observed. Similar findings (i.e., relatively limited transcriptional changes) have been reported in other *Arabidopsis* HMTase mutants (19, 27, 28). Collectively, these results seem to suggest that, in most cases, histone modifications such as H3K4me and H3K36me may be involved in fine-tuning (rather than dictating) transcription levels. These results also indicate that some genes may be much more sensitive to chromatin changes than others. Finally, in addition to the SET domain, plant SDG2 homologs contain several large blocks of highly conserved residues in the N-terminal region (Fig. 1*B*). In SDG2, these conserved residues are dispensable for the in vitro HMTase activity, indicating that they might be involved in other aspects of SDG2 function, such as mediating protein–protein interactions. Interestingly, with the exception of the SET domain, plant SDG2 homologs do not contain other recognizable domains or share any sequence homology with animal and fungal H3K4 HMTases. It is therefore possible that the mechanisms responsible for H3K4 HMTase recruitment in plants might differ from those in animals and fungi.

Methods

Plant Materials and Growth Conditions. All *Arabidopsis* plants used in this study are of the Columbia (Col-0) accession. The *sdg2-1* (SALK_021008) and *sdg2-2* (CS852810, WiscDsLox361D10) were obtained from the *Arabidopsis* Biological Research Center. Plants were grown at 22 °C under continuous light except those used in the flowering time experiments, which were grown under continuous light, long day (16 h light), or short day (8 h light).

Sequence and Phylogenetic Analysis. Blastp and tBlastn searches were performed using the SDG2 amino acid sequence as query against the NCBI nr protein and est databases, respectively. The multiple alignment was generated using CLUSTALW, and the unrooted neighbor-joining tree was constructed using MEGA3.0. Bootstrap values were calculated from 1,000 replicates.

Microarray Analysis and Validation. Wild-type and *sdg2-1* plants were grown side by side, and whole seedlings were harvested 12 d after germination. RNA was extracted using the TRIzol reagent (Invitrogen), cleaned using the RNeasy kit (Qiagen), labeled using the 3′ IVT Express Kit (Affymetrix), and hybridized to the Affymetrix GeneChip *Arabidopsis* ATH1 Genome Array. Array staining, washing, and scanning were performed as previously described (40). Three pairs of biological replicates were included. Data analysis was performed using the GCRMA and LIMMA packages. Raw and processed microarray data have been deposited in Gene Expression Omnibus (accession no. GSE23208).

Real-Time RT-PCR Validation. Reverse transcription was performed using the Transcriptor First Strand cDNA Synthesis Kit (Roche). Real-time PCR was performed using the SYBR Green I Master Mix on a LightCycler 480 (Roche). PCR primers are listed in Table S1.

Protein Purification and HMTase Assays. The 3′ region of the *SDG2* cDNA encoding SDG2C was PCR amplified using the primers 5′-cccgggaGGTCGATCACAAGACTTACGC and 5′-gtcgacCTAACTATCCCATGTCGCTTG, and cloned into the pGEX-6P-1 vector (GE Healthcare). GST-SDG2C fusion protein was expressed in *E. coli* [BL21(DE3); Invitrogen] and purified using the GST Bind Purification Kit (EMD Chemicals). GST-$H3_{1-57}$ as well as the K4R, K9R, and K27R mutant constructs were described previously (41). The K36R and H3K9,27,36R GST-$H3_{1-57}$ constructs were generated using site-directed mutagenesis. HMTase assays were performed as previously described using SAM (Perkin-Elmer) (33). After the HMTase assay, the reaction mix was separated by SDS/PAGE gel electrophoresis, dried, and exposed to films for 6–8 h. Alternatively, the HMTase reaction was performed using unlabeled SAM (NEB), and the products were analyzed by Western blot analysis (see below).

Western Blot and ChIP. Western blot analyses were performed as previously described (42). The antibodies used here are as follows: ab1791 (anti-H3; Abcam), ab8895 (anti-H3K4me1; Abcam), ab32356 (anti-H3K4me2; Abcam),

GENETICS

04-745 (anti-H3K4me3; Millipore), 9E10 (anti-Myc; Covance), and B-14 (anti-GST; Santa Cruz Biotechnology). H3 and H3K4me3 ChIP experiments were performed as previously described (5), using the ab1791 (anti-H3; Abcam) and 07-473 (anti-H3K4me3; Millipore) antibodies, respectively. Real-time PCR primer sequences are listed in Table S2.

ACKNOWLEDGMENTS. We thank Drs. Y. Tamada and R. Amasino (University of Wisconsin-Madison, Madison, WI) for providing the *sdg2-2* seeds. J.A.L. was supported by the National Institutes of Health National Research Service Award 5F32GM820453. Research in the Zhang laboratory was supported by National Science Foundation Grant 0960425.

1. Zhang K, Sridhar VV, Zhu J, Kapoor A, Zhu JK (2007) Distinctive core histone post-translational modification patterns in *Arabidopsis thaliana*. *PLoS ONE* 2:e1210.
2. Johnson L, et al. (2004) Mass spectrometry analysis of *Arabidopsis* histone H3 reveals distinct combinations of post-translational modifications. *Nucleic Acids Res* 32: 6511–6518.
3. Lippman Z, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430:471–476.
4. Turck F, et al. (2007) *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet* 3:e86.
5. Zhang X, et al. (2007) Whole-genome analysis of histone H3 lysine 27 trimethylation in *Arabidopsis*. *PLoS Biol* 5:e129.
6. Bernatavichute YV, Zhang X, Cokus S, Pellegrini M, Jacobsen SE (2008) Genome-wide association of histone H3 lysine nine methylation with CHG DNA methylation in *Arabidopsis thaliana*. *PLoS ONE* 3:e3156.
7. Oh S, Park S, van Nocker S (2008) Genic and global functions for Paf1C in chromatin modification and gene expression in *Arabidopsis*. *PLoS Genet* 4:e1000077.
8. Zhang X (2008) The epigenetic landscape of plants. *Science* 320:489–492.
9. Zhang X, Bernatavichute YV, Cokus S, Pellegrini M, Jacobsen SE (2009) Genome-wide analysis of mono-, di- and trimethylation of histone H3 lysine 4 in *Arabidopsis thaliana*. *Genome Biol* 10:R62.
10. Charron JB, He H, Elling AA, Deng XW (2009) Dynamic landscapes of four histone modifications during deetiolation in *Arabidopsis*. *Plant Cell* 21:3732–3748.
11. Jacob Y, et al. (2010) Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. *Nature* 466:987–991.
12. Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB (1998) Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science* 280:446–450.
13. Alvarez-Venegas R, et al. (2003) ATX-1, an *Arabidopsis* homolog of trithorax, activates flower homeotic genes. *Curr Biol* 13:627–637.
14. Chanvivattana Y, et al. (2004) Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. *Development* 131:5263–5276.
15. Zhao Z, Yu Y, Meyer D, Wu C, Shen WH (2005) Prevention of early flowering by expression of FLOWERING LOCUS C requires methylation of histone H3 K36. *Nat Cell Biol* 7:1256–1260.
16. Kim SY, et al. (2005) Establishment of the vernalization-responsive, winter-annual habit in *Arabidopsis* requires a putative histone H3 methyl transferase. *Plant Cell* 17: 3301–3310.
17. Raynaud C, et al. (2006) Two cell-cycle regulated SET-domain proteins interact with proliferating cell nuclear antigen (PCNA) in *Arabidopsis*. *Plant J* 47:395–407.
18. Johnson LM, Law JA, Khattar A, Henderson IR, Jacobsen SE (2008) SRA-domain proteins required for DRM2-mediated de novo DNA methylation. *PLoS Genet* 4: e1000280.
19. Xu L, et al. (2008) Di- and tri- but not monomethylation on histone H3 lysine 36 marks active transcription of genes involved in flowering time regulation and other processes in *Arabidopsis thaliana*. *Mol Cell Biol* 28:1348–1360.
20. Thorstensen T, et al. (2008) The *Arabidopsis* SET-domain protein ASHR3 is involved in stamen development and interacts with the bHLH transcription factor ABORTED MICROSPORES (AMS). *Plant Mol Biol* 66:47–59.
21. Dong G, Ma DP, Li J (2008) The histone methyltransferase SDG8 regulates shoot branching in *Arabidopsis*. *Biochem Biophys Res Commun* 373:659–664.
22. Pien S, et al. (2008) ARABIDOPSIS TRITHORAX1 dynamically regulates FLOWERING LOCUS C activation via histone 3 lysine 4 trimethylation. *Plant Cell* 20:580–588.
23. Cazzonelli CI, et al. (2009) Regulation of carotenoid composition and shoot branching in *Arabidopsis* by a chromatin modifying histone methyltransferase, SDG8. *Plant Cell* 21:39–53.
24. Jacob Y, et al. (2009) ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. *Nat Struct Mol Biol* 16:763–768.
25. Berr A, et al. (2009) SET DOMAIN GROUP25 encodes a histone methyltransferase and is involved in FLOWERING LOCUS C activation and repression of flowering. *Plant Physiol* 151:1476–1485.
26. Tamada Y, Yun JY, Woo SC, Amasino RM (2009) ARABIDOPSIS TRITHORAX-RELATED7 is required for methylation of lysine 4 of histone H3 and for transcriptional activation of FLOWERING LOCUS C. *Plant Cell* 21:3257–3269.
27. Alvarez-Venegas R, et al. (2006) The *Arabidopsis* homolog of trithorax, ATX1, binds phosphatidylinositol 5-phosphate, and the two regulate a common set of target genes. *Proc Natl Acad Sci USA* 103:6049–6054.
28. Saleh A, et al. (2008) The highly similar *Arabidopsis* homologs of trithorax ATX1 and ATX2 encode proteins with divergent biochemical functions. *Plant Cell* 20:568–579.
29. Baumbusch LO, et al. (2001) The *Arabidopsis thaliana* genome contains at least 29 active genes encoding SET domain proteins that can be assigned to four evolutionarily conserved classes. *Nucleic Acids Res* 29:4319–4333.
30. Alvarez-Venegas R, Avramova Z (2002) SET-domain proteins of the Su(var)3-9, E(z) and trithorax families. *Gene* 285:25–37.
31. Springer NM, et al. (2003) Comparative analysis of SET domain proteins in maize and *Arabidopsis* reveals multiple duplications preceding the divergence of monocots and dicots. *Plant Physiol* 132:907–925.
32. Zhao Z, Shen WH (2004) Plants contain a high number of proteins showing sequence similarity to the animal SUV39H family of histone methyltransferases. *Ann N Y Acad Sci* 1030:661–669.
33. Jackson JP, Lindroth AM, Cao X, Jacobsen SE (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416:556–560.
34. Lindroth AM, et al. (2004) Dual histone H3 methylation marks at lysines 9 and 27 required for interaction with CHROMOMETHYLASE3. *EMBO J* 23:4286–4296.
35. Alvarez-Venegas R, Avramova Z (2005) Methylation patterns of histone H3 Lys 4, Lys 9 and Lys 27 in transcriptionally active and inactive *Arabidopsis* genes and in atx1 mutants. *Nucleic Acids Res* 33:5199–5207.
36. Schmid M, et al. (2005) A gene expression map of *Arabidopsis thaliana* development. *Nat Genet* 37:501–506.
37. Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A (2000) Expression and parent-of-origin effects for *FIS2, MEA*, and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. *Proc Natl Acad Sci USA* 97:10637–10642.
38. Ingouff M, et al. (2009) The two male gametes share equal ability to fertilize the egg cell in *Arabidopsis thaliana*. *Curr Biol* 19:R19–R20.
39. Collins RE, et al. (2005) In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J Biol Chem* 280:5563–5570.
40. Zhang X, et al. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126:1189–1201.
41. Tachibana M, Sugimoto K, Fukushima T, Shinkai Y (2001) Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J Biol Chem* 276: 25309–25317.
42. Jackson JP, et al. (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in *Arabidopsis thaliana*. *Chromosoma* 112:308–315.