Tissue-Specific Differences in Cytosine Methylation and their Association with Differential Gene Expression in *Sorghum bicolor*¹[W]

Meishan Zhang, Chunming Xu, Diter von Wettstein, and Bao Liu*

Key Laboratory of Molecular Epigenetics of MOE and Institute of Genetics & Cytology, Northeast Normal University, Changchun 130024, China (M.Z., C.X., B. L.); Department of Agronomy, Jilin Agricultural University, Changchun 130118, China (M.Z.); and Department of Crop and Soil Sciences and School of Molecular Biology, Washington State University, Pullman WA 99164-6420, USA (D.W.)

Running title: Cytosine Methylation Dynamics across Sorghum tissues

---------------------------------

¹This work was supported by the National Natural Science Foundation of China (No. 30870198 and 30870178), the Program of Introducing Talents of Discipline to Universities in China (B07017), and the US National Institutes of Health grant R01 GM 80749.

*Corresponding author; e-mail baoliu@nenu.edu.cn

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is:

Bao Liu (baoliu@nenu.edu.cn).

[W]The online version of this article contains Web-only data.

www.plantphysiol.org/xxx
It has been well-established that DNA cytosine methylation plays essential regulatory roles in imprinting gene expression in endosperm, and hence normal embryonic development, in the model plant *Arabidopsis thaliana*. Nonetheless, the developmental role of this epigenetic marker in cereal crops remains largely unexplored. Here, we report for *Sorghum bicolor* differences in relative cytosine methylation levels and patterns at 5′-CCGG sites in seven tissues (endosperm, embryo, leaf, root, young-inflorescence, anther and ovary), and characterize a set of tissue-specific differentially methylated regions (TDMRs). We found that the most enriched TDMRs in sorghum are specific for the endosperm and are generated concomitantly but imbalanced by decrease vs. increase in cytosine methylation at multiple 5′-CCGG sites across the genome. This leads to more extensive demethylation in the endosperm than in other tissues, where TDMRs are mainly tissue-nonspecific rather than specific to a particular tissue. Accordingly, relative to endosperm, the other six tissues showed grossly similar levels though distinct patterns of cytosine methylation, presumably as a result of a similar extent of concomitant decrease vs. increase in cytosine methylation that occurred at variable genomic loci. All four tested TDMRs were validated by bisulfite genomic sequencing. Diverse sequences were found to underlie the TDMRs, including those encoding various known-function or predicted proteins, transposable elements (TEs), and those bearing homology to putative imprinted genes in *Zea mays*. We further found that the expression pattern of at least some genic TDMRs was correlated with its tissue-specific methylation state, implicating a developmental role of DNA methylation in regulating tissue-specific or -preferential gene expression in sorghum.

Covalent modification of DNA by 5-methylcytosines is a relatively stable (inheritable) epigenetic marker existing in a vast range of eukaryotic organisms, and particularly abundant in flowering plants. The primary biological function of cytosine methylation is proposed to serve as a genome defense device for silencing transposable elements (TEs) and hence maintaining transgenerational genome integrity (Yoder et al., 1997; Cao et al., 2000; Cao and Jacobsen 2002a,b; Gong et al., 2002; Cao et al., 2003; Kato et al., 2003; Chan et al., 2005; Ding et al., 2007; Berdasco et al., 2008; Suzuki and Bird 2008). This has been robustly supported by large-scale genomic and epigenomic sequencing (Schöb and Grossniklaus 2006; Zhang et al., 2006; Yazaki et al., 2007; Zilberman et al., 2007; Cokus et al., 2008; Lister et al., 2008; Zhang 2008; Gehring et al., 2009; Hsieh et al., 2009; Zemach et al., 2010). Although programmed dynamics of cytosine methylation was known to occur during mammalian development, the developmental role of this epigenetic tool in plants had remained controversial until recently (Schöb and Grossniklaus 2006). Experimental evidence primarily gained in the model plant *Arabidopsis thaliana* but also in barley has unequivocally established the essentiality of cytosine methylation dynamics for normal development by playing critical roles at least during
specific developmental stages (Xiao et al., 2006; Berdasco et al., 2008; Fitzgerald et al., 2008; von Wettstein 2009). For example, similar to the situation in mammals, cytosine methylation mechanistically regulates imprinted gene expression in endosperm and is essential for normal embryonic development in Arabidopsis and in other plants (Sørensen 1992; Spielman et al., 2001; Danilevskaya et al., 2003; Walbot and Evans 2003; Gehring et al., 2004; Kinoshita et al., 2004; Autran et al., 2005; Jullien et al., 2006; Scott and Spielman 2006; Haun et al., 2007; Hermon et al., 2007; Huh et al., 2007).

As in mammals, the major mechanism by which cytosine methylation ramifies plant development is via its regulatory dictation on stable but reversible temporal and spatial gene expression, e.g., through inhibiting transcription initiation by tissue-specific promoter hypermethylation (Sørensen 1992; Sørensen et al., 1996; Cocciolone et al., 2001; Kloti et al., 2002; Sekhon et al., 2007). Indeed, it has further been established in both Arabidopsis and Zea mays that the asymmetric expression of alleles for a given imprinted gene in the central cell during female gametogenesis is accomplished through targeted demethylation of the maternal allele, principally by the 5-methylcytosine DNA-demethylating glycosylase DEMETER (DME), followed by faithful mitotic maintenance of the deprived methylation state of the maternal genome after fertilization. The paternal allele (from sperm) remains methylated throughout male gametogenesis and following fertilization (Walbot and Evans 2003; Gehring et al., 2004; Kinoshita et al., 2004; Jullien et al., 2006; Huh et al., 2007). Thus, identification and characterization of tissue-specific differentially methylated regions (TDMRs) represent an important initial step towards understanding several important related questions, such as, to what extent cytosine methylation varies across plant development, and how TDMRs are established and maintained.

In addition to the recently uncovered extensive, global demethylation in the maternal endosperm-genome of Arabidopsis (Gehring et al., 2009; Hsieh et al., 2009) and rice (Zemach et al., 2010), a large number of endosperm-specific TDMRs were also identified in Z. mays and proposed to account for a 13% reduction of cytosine methylation of the genome in the maize
endosperm when compared to that of the embryo and seedling (Lauria et al., 2004). *Sorghum bicolor* (sorghum) is a close relative of *Z. mays* but with a much smaller genome-size, as well as an important agricultural crop bearing superior tolerance to arid and salinity conditions. Nonetheless, little information is known regarding the occurrence, extent, distribution, and function of cytosine DNA methylation in this important cereal crop plant. The aim of this study was to explore the differences in the level and pattern of cytosine methylation in several representative tissue-types of sorghum, the presence and distribution of TDMRs, and relevance of TDMRs to tissue-specific or -preferential gene expression.

**RESULTS**

**Differences in Relative Levels of Cytosine Methylation among Sorghum Tissues**

Seven tissues (endosperm, embryo, leaf, root, young-inflorescence, anther and ovary) from two sorghum pure-line genotypes and five tissues (endosperm, embryo, leaf, root and young-inflorescence) from a pair of reciprocal F$_1$ hybrids derived from the two pure-lines were analyzed by the methylation-sensitive amplified polymorphism (MSAP) marker. MSAP is a modified version of the amplified fragment length polymorphism (AFLP) by incorporating methylation-sensitive restriction enzymes, *Hpa*II and *Msp*I, which are a pair of isoschizomers that recognize the same tetranucleotide sequence 5´-CCGG (a predominant site for cytosine methylation modification in eukaryotes), but have differential sensitivities to methylation of the cytosine bases (for detailed rationale see *Methods* and Fig. S1). Thus, the marker is an efficient, though gross method to reveal genome-wide relative DNA methylation levels and patterns of randomly sampled 5´-CCGG sites (Dong et al., 2006; Madlung et al., 2005). By using 20 pairs of *Eco*RI+*Hpa*II/*Msp*I selective primer combinations (Table S1), a total of 1,091 clear and reproducible bands (sites) were scored, and polymorphism in the banding patterns between the tissues for a given genotype was detected (e.g., Fig. 1). Tabulating the data according to the methylation state (unmethylated vs. methylated at CG or hemi-methylated at CHG, respectively) of the scored MSAP bands enabled the estimation of the relative total methylation levels at the 5´-CCGG sites of the seven sorghum tissues to be in the range of 24.1-28.2% and composed of
15.1-19.6% $mCG$ and 7.2-9.0% $mCHG$ (Fig. 2). The most striking observation is the more extensive demethylation $mCG$ in endosperm than in the other six tissues (e.g. Fig. 1). Accordingly, the relative methylation level ($mCG + mCHG$) of endosperm is significantly lower than that of the other tissues studied (Fig. 2; $P<0.01$). The $mCHG$ of endosperm is also discernibly lower than in the other six tissues, but the difference is not statistically significant based on the scored MSAP data (Fig. 2). There is no significant difference among the other six tissues for the relative levels of all three types of methylation, $mCG$, $mCHG$ and total ($mCG + mCHG$) ($P>0.05$). Between-genotype difference in the relative methylation level is insignificant in all possible pairwise comparisons for a given tissue; however, the two reciprocal $F_1$ hybrids appeared to have lower relative levels of CG methylation (and hence of total) in each of the five tissues relative to their parental inbred pure-lines, but again the difference is statistically insignificant ($P>0.05$), indicating largely stable meiotic inheritance of the pure-line parental methylation states to their $F_1$ hybrids (Fig. 2).

**Difference in Cytosine Methylation Patterns among Sorghum Tissues**

Differential sensitivity to cytosine methylation modifications at the 5'-CCGG sites by the pair of isoschizomers ($HapII/MspI$) enables unequivocal categorization of the tissue-specific MSAP patterns into four major types: CG hypermethylation ($mCG$), CG hypomethylation, CHG hypermethylhation ($mCHG$), and CHG hypomethylation (Zhang et al., 2009). To compare difference in cytosine methylation patterns among the studied sorghum tissues, we first set aside one tissue (embryo) as a reference, and quantified the difference of the four methylation patterns for the other six tissues relative to embryo. All six tissues showed differences in both hyper- and hypomethylation relative to the reference embryo tissue. This points to the existence of tissue-specific differentially methylated regions (TDMRs) in the sorghum genome, as detailed below. Relative to embryo, five of the six tissues, except endosperm, showed lower frequencies ranging from 0.1% to 1.7% of variation in hyper- and hypomethylation for both CG and CHG sites to a similar extent (Fig. 3). This may largely explain the gross similarity in the net relative methylation levels among these tissues (Fig. 2). In contrast, for endosperm, hypomethylation of
both CG and CHG (ranging from 2.6% to 7.5%) markedly exceeds hypermethylation (ranging from 1.1% to 1.4%) by factors 2 and 7, respectively (Fig. 3). This may explain the significantly reduced mCG level and clear though statistically insignificant mCHG level in endosperm relative to the other six tissues (Fig. 2). To a large extent, the methylation pattern and relative level of each tissue in the reciprocal F₁ hybrids are additive of their pure-line parents (Figs. 1, 2 and 3), consistent with Mendelian transmission genetics.

To gain additional insights into the tissue-specific difference in cytosine methylation patterns of mCG and mCHG, we performed all possible pairwise comparisons of the seven tissues for each of the four genotypes. We obtained the following results: (1) As described above, relative to the other six tissues, endosperm showed predominantly CG hypomethylation (frequencies of 7-8%), followed by CHG hypomethylation (frequencies of 2-4%); nonetheless, hypermethylation of both CG and CHG (at frequencies of 0.5-2%) is also evident in endosperm relative to the other six tissues (Fig. 4a). (2) Relative to embryo, leaf showed the least difference in both mCG and mCHG, followed by young-inflorescence and the two gametophyte tissues (anther and ovary), with root showing the largest methylation difference from embryo (Fig. 4b). (3) Relative to leaf, young-inflorescence showed the least difference in both mCG and mCHG, followed by root, with the two gametophyte tissues (anther and ovary) showing the largest methylation difference from leaf (Fig. 4c). (4) Relative to root, young-inflorescence showed less difference in both mCG and mCHG than the two gametophyte tissues (anther and ovary) (Fig. 4d). (5) The two gametophyte tissues showed very little difference in both mCG and mCHG from each other (Fig. 4e). (6) In all possible pairwise comparisons described above, genotypic differences among the four lines used (two pure-lines and their reciprocal F₁ hybrids) was nonexistent or negligible (Fig. 4), indicating again conservation and stable Mendelian inheritance (from parental pure-lines to F₁ hybrids) of the TDMRs in sorghum.

Isolation and Characterization of Tissue-Specific Differentially Methylated Regions (TDMRs) in Sorghum
Based on nature of the tissue-specific polymorphism of the MSAP banding patterns among the seven or five tissues being compared for a given genotype, we scored, isolated and sequenced a subset of the tissue-specific differentially methylated regions (TDMRs). These can be categorized into two groups: group one is tissue-specific referring to those that are unmethylated in only one of the seven tissues, and group two is tissue-nonspecific referring to those that are unmethylated in two to six out of the seven tissues. Because homoplasy was not a concern for these closely related sorghum genotypes belonging to the same species, we isolated the TDMRs only from one genotype at a given position in the MSAP profiles, although they were scored independently for each genotype (Table I). Actually, we found genotype-specific TDMRs of both tissue-specific and -nonspecific types to be very rare, and which were only detected in one or both of the F₁ hybrids (Table I). We quantified the relative distribution of the identified TDMRs among the seven tissues compared (Fig. 5). We found that the majority of the TDMRs were endosperm-specific demethylation for either mCG or mCHG though a greater proportion for mCG (52.1-62.0%) than for mCHG (33.8-44.8%), the rest being those that are demethylated in two to six tissues but methylated in other tissues, i.e., tissue-nonspecific TDMRs (2.3% to 12.5% for mCG, and 5.2% to 15.9% for mCHG). Those that are specific to each of the rest six tissues apart from endosperm altogether constitute only a very small proportion (0 to 5.0% for mCG, and 0 to 2.9% for mCHG) (Fig. 5; Table S3).

Sequencing of the TDMRs indicated that some are redundant, and in total we obtained 58 distinct TDMRs (Table II). Of these, 42 are tissue-specific and 16 are tissue-nonspecific (Table S3). A blast X analysis of these TDMRs indicated that they are functionally diverse (Table II; Table S3). Specifically, 17 bear significant homology to genes coding for known-function proteins (Table II; Table S3), which include seven metabolism-related proteins (TDMRs 9, 12, 27, 48, 52, 40 and 56), four endosperm proteins (TDMRs 50, 43, 44 and 45), a senescence-associated protein (TDMR42), a nuclear ribonucleoprotein (TDMR25), a disease-related protein (TDMR18), and a helicase-like protein (TDMR35) (Table S3). Notably, of these 17 TDMRs, two (45 and 50; both being endosperm-specific) are homologous to previously characterized putative imprinting genes in Z. mays, and which were shown to exhibit
endosperm-preference or -specific expression patterns (Danilevskaya et al., 2003; Guo et al., 2003). Six TDMRs (13, 28, 51, 53, 54 and 55) showed meaningful homology to TE (retrotransposon)-related proteins. Nine TDMRs showed meaningful homology to hypothetical proteins. The rest 26 TDMRs showed either no homology or homology to anonymous sorghum cDNAs of various sources (Table II and Table S3).

**Validation of TDMRs in Sorghum by Bisulfite Genomic Sequencing**

We selected four TDMRs as representatives to perform bisulfite genomic sequencing for each of the seven tissues of genotype pure-line A (as these TDMRs are common among all four genotypes in the MSAP profiles). These four TDMRs included a low-copy number TE (a retrotransposon fragment, TDMR28), a normal genic (kinase) fragment (TDMR40), and the two fragments (TDMR45 and TDMR50) bearing homology to putative imprinted genes in *Z. mays* (Table S3). Of these four TDMRs, three (TDMR28, TDMR45 and TDMR50) are endosperm-specific (unmethylated only in endosperm and methylated in all the rest six studied tissues), and one (TDMR40) is tissue-nonspecific (unmethylated in endosperm and root, methylated in the rest five tissues) (Table S3). We found that, indeed, for all three endosperm-specific TDMRs, the methylation levels at both CG and CHG sites were substantially (by 66.5 to 93%) reduced in endosperm relative to the rest six tissues being compared, although among which moderate difference (from 5 to 20%) in methylation was also detected (Fig. 6; Fig. S2; Table S3). For the tissue-nonspecific TDMR (TDMR40), endosperm and root indeed showed reduced CG methylation level by ca. 20% than the other five tissues, but this is not the case for CHG methylation which exhibited fluctuation across the tissues without a clear trend (Fig. 6). Methylation of asymmetric CHH cytosines for all four investigated TDMRs was very low (in most cases <5%), and did not manifest substantial changes among the tissues in three of the four analyzed TDMRs; nonetheless, in one of the TDMRs (TDMR45), CHH methylation level of embryo (~12%) was conspicuously higher than the rest tissues (ranging from 1 to 4%) (Fig. 6). Together, authenticity of all four investigated TDMRs was validated by bisulfite genomic sequencing, suggesting that the great majority, if
not all, of the sorghum TDMRs uncovered by the MSAP analysis are *bona fide* tissue-differential methylated loci. However, we can not rule out the possibility that a small proportion of these TDMRs identified by MSAP may actually represent incidental capture of the methylation state of minority cells of a given tissue due to the involvement of two-round PCR amplifications in the technique.

**Tissue-specific or -preferential Expression of Genes Associated with TDMRs in Sorghum**

We selected nine TDMRs (28, 25, 40, 42, 44, 48, 45, 43 and 50) bearing homology to either protein-encoding genes or TEs (Table S3), including the bisulfite-sequenced four (Fig. 6), as representatives for gene expression analysis across the seven tissues, i.e., from the two pure-lines and five of the pair reciprocal F₁ hybrids. We measured the relative steady-state transcript abundance with semi-quantitative RT-PCR amplifications (Fig. 7) on three batches of RNAs independently isolated from different collections for each of the tissues. Note that the results for each sequence were completely validated by real-time q-RT PCR analysis (data not shown). We found that whereas the retrotransposon (TDMR28) showed no expression in any of the seven tissues including endosperm, the rest eight genes showed clear expression in at least one of the studied tissues (Fig. 7). Absence of expression of TDMR28 occurred in spite of substantial hypomethylation at its CG sites (Fig. 6). Specifically, we found that (1) the three tissue-specific or -nonspecific TDMRs, 25, 40 and 42, which were differentially methylated at CG or both CG and CHG based on the MSAP profile (Table S3), were expressed in all the seven tissues investigated, with only modest difference in expression level among the tissues; (2) the two tissue-specific TDMRs, 44 and 48, which were differentially methylated at CG or CHG based on the MSAP profile (Table S3), were also expressed in all the seven tissues, but with significant difference in expression level among the tissues; (3) the endosperm-specific TDMR (45), which was unmethylated only in endosperm at CG according to MSAP (Table S3), but actually also at both CG and CHG sites based on bisulfite sequencing (Fig. 6), was indeed expressed at the highest level in endosperm, at a moderate level in young-inflorescence, but it was also expressed, though at much reduced levels, in embryo, anther and ovary, and
completely silenced in leaf and root; (5) the endosperm-specific TDMR (43), which was unmethylated only in endosperm at both CG and CHG based on MSAP (Table S3), was unexpectedly expressed only in anther and silenced in all the rest six tissues including endosperm; (6) the endosperm-specific TDMR (50), which was unmethylated only in endosperm at CG and CHG based on MSAP (Table S3), as verified by bisulfite sequencing (Fig. 6), was indeed expressed only in endosperm and silenced in all the rest six tissues. Taken together, it appeared that some, but not all, of the TDMR-associated genes showed variable expression patterns among the seven sorghum tissues, which are in accord with their differential methylation states in the tissues. The most clear-cut cases being the two endosperm-specific TDMRs (45 and 50), which were previously shown in maize as putative imprinted genes (Danilevskaya et al., 2003; Guo et al., 2003), were indeed preferentially or specifically expressed only in the sorghum endosperm. Nonetheless, we note that the expression patterns of some of the TDMRs among the tissues are apparently incongruent with their methylation states based on the MSAP profiles, particularly of those that showed differential CG methylation (e.g., TDMR42, TDMR44 and TDMR48), suggesting that CHG methylation likely plays a more important role for orchestrating tissue-specific or -preferential expression in sorghum. The TDMR that showed the most incompatible relationship between its methylation state and expression is TDMR43, as it was expressed only in anther and silenced in all the rest six tissues in spite of the fact that it is unmethylated in endosperm but methylated in all the rest six tissues including anther (Table S3).

DISCUSSION

Difference in Relative Level and Pattern of Cytosine Methylation among Sorghum Tissues

We have shown in this study that amongst the seven sorghum tissue types (endosperm, embryo, leaf, root, young-inflorescence, anther and ovary) analyzed by the MSAP marker, endosperm is the only tissue that exhibits substantial reduction in the relative gross level of $^m$CG and to a less extent also $^m$CHG relative to the other six tissues -- which however showed little difference among themselves in the levels of both $^m$CG and $^m$CHG. This result is in general
agreement with the recently depicted bisulfite sequencing-based methylomes of various rice tissues (Zemach et al., 2010). In particular, similar to the situation in rice (Zemach et al., 2010), we also detected a general increase of CHG methylation in embryo relative to the other five tissues (except endosperm) as far as the 5\'-CCGG sites are concerned (compare Figs. 4b with c, d and e), implicating the involvement of small RNA-guided mechanisms (Cao et al., 2003; Tang, 2005; Mathieu et al. 2007; Matzke et al., 2009). Further comparison of methylation patterns between the sorghum tissues, either with reference to one specific tissue (embryo) or in a pairwise manner, revealed that the reduction of mCG and mCHG levels in endosperm was not due to a unidirectional demethylation of mCG and mCHG. Rather, relative to the other six tissues, the net loss of cytosine methylation in endosperm resulted from imbalanced, decrease- vs. increase in methylation, with the former being much greater in magnitude and spectrum than the latter (Fig. 2). Similarly, the pairwise comparisons revealed that each tissue is distinct in terms of methylation patterns, because the locus-specific hypo- and hypermethylation changes occurred at variable 5\'-CCGG sites across the genome. This is the case in spite of the fact that the other six tissues (apart from endosperm) are very similar in the relative levels of both mCG and mCHG due to a largely balanced decrease- vs. increase in methylation in one tissue vs. another. Notably, this feature holds true for all four sorghum genotypes studied, i.e., the two pure-lines and a pair of reciprocal F1 hybrids derived from them, indicating that the difference in the relative level and pattern of cytosine methylation among the sorghum tissues is highly conserved across genotypes, as well as meiotically inheritable from parental pure-lines to F1 hybrids in a Mendelian manner.

That the differential cytosine methylation patterns among the sorghum tissues are largely due to balanced or imbalanced concomitant hyper- and hypomethylation changes in one tissue vs. another has not been reported previously in plants. Nevertheless, it bears remarkable similarity to the recently documented trend of methylation alteration in human colon cancer tissue relative to adjacent normal tissues, wherein similar degree of hypo- and hypermethylation were detected in the former (Irizarry et al., 2009). The endosperm vs. embryo comparison of cytosine methylation patterns in the MSAP profiles of sorghum is also grossly compatible with
characteristics of the recently established genomic methylation landscapes (methylomes) of the *Arabidopsis* and rice endosperm and embryo. These studies showed that the genomewide demethylation of the genome in the endosperm is also accompanied by localized hypermethylation of CHG and CHH but not of CG, primarily due to a burst of small RNAs in the endosperm (Hsieh et al., 2009; Zemach et al., 2010). Our results revealed additionally that compared with the embryo and other tissues, CG-hypermethylation occurred in the sorghum endosperm at least in 1-2% of the sampled 5'-CCGG sites (Fig. 4a). This is novel but also compatible with a small-RNA-guided mechanism like RdDM being invoked, which targets cytosines of all sequence context (CG, CHG and CHH) for de novo methylation (Cao et al., 2003; Tang, 2005; Mathieu et al. 2007; Matzke et al., 2009). However, as our results are only based on the analysis of the 5'-CCGG sites randomly sampled across the genome, it remains to be tested whether the conclusion can be extrapolated to all the CG sites across the whole sorghum endosperm (maternal) genome.

**Characteristics of TDMRs in Sorghum**

The concomitant occurrence of hypo- and hypermethylation changes at variable 5'-CCGG loci in one tissue relative to another or others enables in sorghum identification of TDMRs belonging to two categories, tissue-specific and tissue-nonspecific. Consistent with the findings in mammals (Song et al., 2005) and other plants, e.g., *Arabidopsis* (Kinoshita et al., 2004; Ruiz-Garcia et al., 2005), maize (Lauria et al., 2004) and rice (Zemach et al., 2010), we also found in sorghum that the most enriched TDMRs are those that are specifically hypomethylated in endosperm (or placenta in mammalian animals), followed by those that are tissue-nonspecific. The later type are hypomethylated in two or more tissues but hypermethylated in at least one of the tissues being compared with those that are specific to any one of the tissues other than endosperm being very rare or probably non-existent.

Homology analysis of a subset (58) of isolated TDMRs indicated that they represent diverse sequences, and a substantial proportion bears significant homology to protein-coding genes of various functional categories. Probably more than incidental, two of the characterized
endosperm-specific TDMRs are homologous to previously characterized putative imprinted genes in *Z. mays*, and six TDMRs are homologous to TE (retrotransposon)-related genes. These sorghum TDMRs are again highly conserved across the four genotypes studied, as well as share characteristics of TDMRs uncovered in maize (Lauria et al., 2004) and *Arabidopsis* (Kinoshita et al., 2004; Ruiz-Garcia et al., 2005). Together, these results support the recent proposal that demethylation in the maternal endosperm genome is a universal process affecting numerous genomic loci rather than being specifically targeted to the imprinted genes; instead, the various TEs, which constitute the bulk of most plant genomes, are likely the primary targets for genomewide demethylation in endosperm, with loss of methylation in the imprinted genes being accomplished as byproducts of the process (Zemach et al., 2010).

Notably, a recent study has revealed significant difference in the \textsuperscript{5}CG demethylation patterns in rice endosperm from that in *Arabidopsis* (Zemach et al., 2010). This study has documented that, relative to embryo and other tissues, the rice endosperm manifests localized, rather than even, \textsuperscript{5}CG demethylation across the genome as found in *Arabidopsis* (Gehring et al., 2009; Hsieh et al., 2009); in contrast, \textsuperscript{5}CHG and \textsuperscript{5}CHH demethylation did not show any difference between endosperm of the two species, as both exhibiting genomewide even demethylation (Zemach et al., 2010). Our results, being based on the MSAP marker, did not allow a precise genomewide comparison including all the cytosine residues; however, based on the randomly sampled 5'-CCGG sites, it is clear that the significant decrease in the relative methylation level of endosperm is primarily due to \textsuperscript{5}CG-demethylation, as there is virtually no significant difference in the CHG methylation level between endosperm and embryo or other tissues (Fig. 2). Nonetheless, our bisulfite sequencing data on the four representative TDMRs actually showed that the reduction of methylation levels at the CHG sites in endosperm relative to the rest six tissues being compared were even higher in magnitude though not in absolute values (due to the intrinsically lower degree of \textsuperscript{5}CHG relative to \textsuperscript{5}CHH) than that of the CG sites (Fig. 6 and Fig. S2). These results imply that the recent documentation in rice that endosperm CHG and CHH demethylation occurs evenly across the genome whereas \textsuperscript{5}CG
demethylation is localized (Zemach et al., 2010) probably also applies to sorghum, which is conceivable given that the two species are both grasses and phylogenetically akin to each other.

**Correlation of TDMRs with Tissue-specific or -preferential Gene Expression in Sorghum**

Although the highly conserved nature (even between species) of TDMRs identified from diverse organisms would have intuitively implied a functional role, the correlation of TDMRs with tissue-specific or -preferential gene expression has been controversial even in mammals in which extensive investigations have been conducted (Walsh and Bestor 1999; Song et al., 2005). Novel insights gained recently both with respect to the high-resolution global methylation landscapes in *Arabidopsis* and high-throughput characterization of a large number of human TDMRs have provided novel insights towards reconciling the discrepant relationships between TDMRs and differential expression of their associated genes. First, because gene-body regions are more likely methylated than their promoters, and methylation of the former regions may not adversely affect transcriptional activity (Zhang et al., 2006; Zilberman et al., 2007), it is logic that some of the TDMRs may not show an expected correlation with transcriptional activity of their associated genes if they are landing on the gene-body regions. Second, a recent study has documented that most TDMRs in normal human tissues, as well as those loci undergoing methylation alteration in cancer cells turned out to be more likely residing in genomic regions outside of the canonical CpG islands, and therefore the expected linear relationship between methylation and gene expression may be obscured by these TDMRs (Irizarry et al., 2009).

In this study, the great majority (19 out of 23) of the isolated sorghum TDMRs that could be assigned to specific regions based on the incompletely annotated sorghum genome (http://www.phytozome.net/search) were found to map to gene-body regions (Table S3), consistent with the methylation landscape of plants (Zhang et al., 2006; Zilberman et al., 2007). However, the steady-state transcript abundance for two of the nine tested genes (TDMR45 and TDMR50, both being imprinted genes) exhibited a clear inverse relationship with their methylation state across the seven studied tissues, suggesting that the inter-tissue methylation difference even within gene-body regions may still play a role in regulating tissue-specific or
-preferential expression, although we cannot rule out the possibility that the promoter regions of these two genes are also demethylated in endosperm.

The mechanism for the establishment and maintenance of TDMRs in sorghum and other non-model plants is currently unknown. It has been established in *Arabidopsis* that the genomewide demethylation in the endosperm-genome is primarily due to hyper-activity of the DNA-demethylating glycosylase DEMETER (DME) in the central cells during female gametogenesis, and followed by faithful maintenance of the hypomethylated state in the maternal genome subsequent to fertilization (Choi et al., 2002; Walbot and Evans 2003; Gehring et al., 2004; Kinoshita et al., 2004; Autran et al., 2005; Huh et al., 2007; Jullien and Berger 2009). Recently, Zemach et al., (2010) have proposed that monocots do not contain DME based on phylogenetic analysis of the nucleotide sequence orthologues, and active demethylation in these plants are due to the action of ROS1 (Gong et al., 2002) and DEMETER-like 3 (DML3) proteins. Nonetheless, an earlier study clearly indicated conservation in rice of the DNA glycosylase domain, cysteine cluster, and lysine-rich region of the *Arabidopsis* DME gene (Choi et al., 2002). Moreover, we were able to isolate a full-length DME gene from barley (von Wettstein 2009) and three DME genes from common wheat; functional analysis showed that these DME genes have clear glycosylase activity (our unpublished data). Therefore, we consider it is still an open possibility that DME proteins may actually exist and be responsible for active demethylation of the maternal endosperm genome in monocots, which warrants further investigations.

**MATERIALS AND METHODS**

**Plant Lines**

Two sorghum pure strains, YN378 (A) and YN406 (B), were originally provided by the Institute of Crops, Jilin Academy of Agricultural Sciences, Changchun, China, and then being maintained in our hands by strict self-pollination for many generations. A pair of reciprocal inter-strain F1 hybrids, designated as AB and BA (the first and second letters denoting maternal
and paternal parents, respectively) were produced by careful manual pollination before the study. Endosperms and embryos were harvested at 15 dpa (days post pollination), leaf tissues were harvested from 9-10th leaf-stage of plants, and young inflorescences were collected at 0.5-1.0 cm in length. Anthers and ovaries were harvested at 2-3 days before flowering (from pure line A and B only). All the above tissues were taken from field-grown plants under normal conditions in accordance with season. To avoid microbial contamination, the root tissue used for DNA isolation was collected from aseptically cultured seedlings in half-strength Murashige and Skoog medium for two weeks.

**MSAP Analysis**

Genomic DNA was isolated from different tissues of a given genotype using the high-salt CTAB method. MSAP (methylation-sensitive amplified polymorphism) was using EcoRI coupled with HpaII and MspI, a pair of isoschizomers that recognize the same restriction site (5'-CCGG) but have different sensitivity to methylation of the cytosines (Fig. S1). HpaII can not cut if either of the cytosines in the double-strand is methylated, whereas MspI cannot cut if the external cytosine is fully- or hemi- (single-strand) methylated. Thus, for a given DNA sample, full methylation of the internal cytosine (marked as \(^\text{mCG}\)), or hemi-methylation of the external cytosine (marked as \(^\text{mCHG}\)), at the assayed 5'-CCGG sites, will be revealed as presence in only one of the enzyme digests and absent from the other. On the other hand, simultaneous hyper- or hypomethylation of both Cs will be revealed as concomitant absence or presence of a band with both enzymes relative to other sample(s) of the same genotype with unmethylated Cs at the 5'-CCGG site in question. In total, one pair of pre-selective primers and 20 pairs of selective primers (Table S1) were used and which allowed scoring of >1,000 5'-CCGG-containing loci. The restriction enzymes EcoRI, HpaII and MspI were purchased from New England Biolabs Inc. (Beverly, Mass.). The amplification products of MSAP were resolved in a 5% denaturing polyacrylamide gel electrophoresis and visualized by silver staining. Only clear and completely reproducible bands in two experiments using independent DNA extractions were scored. Statistical analysis was performed by the \(t\)-test using SPSS statistics.
Identification and Sequencing of TDMRs from the MSAP Profiles

A subset of bands showing methylation difference among various tissues (TDMRs) were eluted from the silver-stained MSAP gels and re-amplified with the appropriate selective primer combinations. Sizes of the PCR products were verified by agarose gel electrophoresis, and then cloned into the pMD18-T vector (Takara Biotech. Inc., Dalian), and sequenced with vector primers by automated sequencing. The Advanced BlastN and BlastX programs at the NCBI website (http://www.ncbi.nlm.nih.gov/) and the sorghum whole-genome sequence at the phytozome website (http://www.phytozome.net/search) were used for homology analysis of the cloned TDMRs.

Bisulfite Genomic Sequencing

Genomic DNA was modified by the EZ DNA Methylation-Gold Kit (Zymo Research, USA) according to the manufacturer’s recommendations. Briefly, 900 μl water, 50 μl M-dissolving Buffer, and 300 μl M-dilution Buffer were added per tube of CT Conversion Reagent prior to use. One hundred and 30 μl of bisulfite-containing CT-Conversion Reagent was added to 0.5-1 μg of DNA in a volume of 20 μl and mixed, and samples were incubated at 98ºC for 10 min, and 64ºC for 2.5 h. Modified DNA was purified by Zymo-Spin IC Column and stored at -20ºC until used. For each PCR, between 1.0 μl and 3.0 μl of bisulfite-treated DNA was used and the PCR products were cloned into the pMD18-T vector (Takara Biotech. Inc., Dalian) and sequenced. More than 20 or 30 clones were sequenced for each experiment. Primers for three unique-copy genic sequences (TDMR40, TDMR45 and TDMR50) and one low-copy retrotransposon (TDMR28) for bisulfite sequencing were designed by the Meth-Primer Program (http://www.urogene.org/methprimer/) and given in Table S2. The methylation levels per sequence motif (CG, CHG, and asymmetric CHH), referred to in Fig. 5 and Fig. S2 as “methylation percentage (%),” were calculated by dividing the number of non-converted cytosines by the total number of cytosine positions within the assay.

Semi-quantitative RT-PCR Analysis for Gene Expression
The protocol was essentially as reported (Liu et al., 2004). Specifically, total RNA was isolated from various tissues of the same set of sorghum plants as used for DNA isolation, from a pair of reciprocal sorghum inter-strain F1 hybrids (AB and BA) and its parental pure lines (A and B), by the Trizol Reagent (Invitrogen), following the manufacturer’s protocol. The RNA was treated with DNaseI (Invitrogen), reverse-transcribed by the Super-Script-TM RNase H-Reverse Transcriptase (Invitrogen), and subjected to RT-PCR analysis using gene-specific primers. The primers for the actin gene, nine sequenced MSAP bands that bear significant homology to annotated genes in sorghum or other plants were designed by the Primer 3 program (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and are given in Table S2. DNA contamination was tested by inclusion of RNAs without RT. The amplicons were visualized by ethidium-bromide staining after electrophoresis through 2% agarose gels. Three batches of independently prepared total RNAs were used as technical replications.

Supplemental Data

The following materials are available in the online version of this article.

**Figure S1.** Schematic representation of the MSAP technique used in this study --EcoRI-AAC and HpaII/MspI-TCT (a1) as an example.

**Figure S2.** Sequence and methylation map for each of the four TDMRs subjected to bisulfite sequencing.

**Table S1.** Sequence of adaptors, pre-amplification primers and selective amplification primer combinations used in the MSAP analysis.

**Table S2.** Primers used for semi-quantitative RT-PCR analysis and bisulfite sequencing.

**Table S3.** Chromosomal location, property, tissue-specific methylation state, putative function, and restriction map of the sequenced sorghum TDMRs.

ACKNOWLEDGMENTS

We are grateful to Dr. Sachin Rustgi and two anonymous reviewers for their valuable comments and suggestions to improve the manuscript.

LITERATURE CITED

version of an Oedipus complex. Curr Opin Plant Biol 8: 19-25


Sørensen MB (1992) Methylation of B-hordein genes in barley endosperm is inversely correlated with gene activity and affected by the regulatory gene Lys3. Proc Natl Acad Sci USA 89: 4119-4123


Figure legend

Figure 1. Examples of MSAP profiles illustrating the various tissue-specific differentially methylated regions (TDMRs) (arrowed) across the seven studied sorghum tissues of a pair of reciprocal inter-strain F1 hybrids (AB and BA) and its parental pure-lines (A and B). Primer combinations used are EcoRI+ACG/HpaII (MspI)+TTA (a), EcoRI+AGG/HpaII (MspI)+TTG (b), EcoRI+AAC/HpaII (MspI)+TCT (c), and EcoRI+ACA/HpaII (MspI)+TTG (d).

Figure 2. Methylation levels of the two major types, mCG and mCHG, calculated based on the MSAP profiles in five tissues of a pair of reciprocal sorghum inter-strain F1 hybrids (AB and BA) and seven tissues in their parental pure-lines (A and B). **denotes statistical significance at 0.01 level based on the Student’s t-test.

Figure 3. Hypo- and hypermethylation at CG or CHG of the randomly sampled 5’-CCGG sites by the MSAP marker in four (a pair of reciprocal F1 hybrids, AB and BA) or six (pure-lines, A and B) sorghum tissues with reference to embryo.

Figure 4. Pairwise comparison of methylation difference at CG or CHG of the randomly sampled 5’-CCGG sites by the MSAP marker among the five (a pair of reciprocal F1 hybrids, AB and BA) or seven (two pure-lines, A and B) sorghum tissues.

Figure 5. Relative distribution of the sorghum tissue-specific differentially methylated regions (TDMRs) based on the scored MSAP profiles of the five (a pair of reciprocal F1 hybrids, AB and BA) or seven (pure-lines, A and B) studied tissues.

Figure 6. Collective cytosine methylation values (in percentage) of CG, CHG and CHH for each of the seven sorghum tissues. The tabulated values are based on the bisulfite sequencing data (given in Figure S2) for each of the four selected TDMRs: a low-copy retrotransposon (TDMR28), a putative sorghum protein kinase gene (TDMR40), a homologue of the Zea mays endosperm NAM-related protein-coding gene (TDMR45), and a homologue of the Zea mays fertilization-independent endosperm protein 1-coding gene (TDMR50).
Table I. Summary of scored TDMRs across the seven sorghum tissues in each of the two pure-line genotypes and a pair of reciprocal F1 hybrids derived from the two pure-lines

<table>
<thead>
<tr>
<th>TDMRs</th>
<th>Endosperm</th>
<th>Embryo</th>
<th>Leaf</th>
<th>Root</th>
<th>Young inflorescence</th>
<th>Anther</th>
<th>Ovary</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Tissue-specific</td>
<td>86</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AB Tissue-specific</td>
<td>91</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tissue-nonspecific</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>34</td>
</tr>
<tr>
<td>BA Tissue-specific</td>
<td>92</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>7</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Tissue-nonspecific</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>B Tissue-specific</td>
<td>79</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
</tbody>
</table>
Table II. Functional categorization of a set of sequenced sorghum TDMRs†

<table>
<thead>
<tr>
<th>Category</th>
<th>Known-function protein</th>
<th>Transposon &amp; retrotransposon</th>
<th>Predicted-protein</th>
<th>No homology or homology to anonymous cDNAs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. TDMRs</td>
<td>17</td>
<td>6</td>
<td>9</td>
<td>26</td>
<td>58</td>
</tr>
<tr>
<td>Percentage (%)</td>
<td>29.3</td>
<td>10.3</td>
<td>15.5</td>
<td>63.1</td>
<td>100</td>
</tr>
</tbody>
</table>

†Detailed information regarding these TDMRs is given in Table S3.
The bar chart shows the methylation level (in %) of CG and CHG sites in various tissues:

- **A:** 1: Endosperm, 2: Embryo, 3: Leaf, 4: Root, 5: Young inflorescence
- **AB:** 1: Anther, 2: Ovary
- **BA:** 1: Endosperm, 2: Embryo, 3: Leaf, 4: Root, 5: Young inflorescence
- **B:** 1: Anther, 2: Ovary

The chart uses different colors to indicate the methylation level of CG (white) and CHG (black) sites.
<table>
<thead>
<tr>
<th></th>
<th>Endosperm</th>
<th>Embryo</th>
<th>Leaf</th>
<th>Root</th>
<th>Young inflorescence</th>
<th>Anther</th>
<th>Ovary</th>
<th>Cycles</th>
<th>bp</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TDMR28</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>367</td>
<td>35</td>
</tr>
<tr>
<td><strong>TDMR25</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>390</td>
<td>26</td>
</tr>
<tr>
<td><strong>TDMR40</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>372</td>
<td>29</td>
</tr>
<tr>
<td><strong>TDMR42</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>392</td>
<td>25</td>
</tr>
<tr>
<td><strong>TDMR44</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>484</td>
<td>25</td>
</tr>
<tr>
<td><strong>TDMR48</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>485</td>
<td>25</td>
</tr>
<tr>
<td><strong>TDMR45</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>485</td>
<td>25</td>
</tr>
<tr>
<td><strong>TDMR43</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>378</td>
<td>28</td>
</tr>
<tr>
<td><strong>TDMR50</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>344</td>
<td>28</td>
</tr>
<tr>
<td><strong>Actin(-RT)</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>352</td>
<td>25</td>
</tr>
<tr>
<td><strong>Actin</strong></td>
<td>A</td>
<td>AB</td>
<td>BAB</td>
<td>A</td>
<td>B</td>
<td>A</td>
<td>B</td>
<td>352</td>
<td>25</td>
</tr>
</tbody>
</table>