Identification of putative *Arabidopsis* DEMETER target genes by GeneChip analysis

Hyonhwa Ohr [a], Anhthu Q. Bui [b], Brandon H. Le [b], Robert L. Fischer [c], Yeonhee Choi [a,*]

[a] *Department of Biological Sciences, Seoul National University, Seoul 151-742, Republic of Korea*
[b] *Department of Molecular, Cell, and Developmental Biology, University of California—Los Angeles, Los Angeles, CA 90095, USA*
[c] *Department of Plant and Microbial Biology, University of California—Berkeley, Berkeley, CA 94720, USA*

Abstract

The *Arabidopsis* DEMETER (DME) DNA glycosylase is required for the maternal allele expression of imprinted Polycomb group (*MEDEA* and *FIS2*) and transcription factor (*FWA*) genes in the endosperm. Expression of DME in the central cell, not in pollen or stamen, establishes gene imprinting by hypomethylating maternal alleles. However, little is known about other genes regulated by DME. To identify putative DME target genes, we generated *CaMV:DME* plants which ectopically express *DME* in pollen and stamens. Comparison of mRNA profiles revealed 94 genes induced by ectopic DME expression in both stamen and pollen. Gene ontology analysis identified three molecular functions enriched in the DME-inducible RNA list: DNA or RNA binding, kinase activity, and transcription factor activity. Semi-quantitative RT-PCR verified the candidate genes identified by GeneChip analysis. The putative target genes identified in this study will provide insights into the regulatory mechanism of DME DNA glycosylase and the functions of DNA demethylation.

Double fertilization is a unique characteristic of flowering plants. Fusion of an egg cell with a sperm cell gives rise to the diploid embryo and the second fusion of a diploid central cell with another sperm cell gives rise to the triploid endosperm. The endosperm is the functional homologue of the mammalian placenta and provides nutrients to the developing embryo.

Gene imprinting refers to the differential allele expression of paternal and maternal alleles depending on parental origin. Imprinting occurs in the placenta of mammals [1] and in the endosperm of plants [2,3]. *MEDEA* (*MEA*) is a Polycomb group (PcG) gene imprinted in *Arabidopsis* [4–6]. The maternal *MEA* allele is expressed while the paternal allele is silenced [7]. There are two mechanisms regulating *MEA*

imprinting. Demethylation by the DEMETER (DME) DNA glycosylase activates expression of the maternal *MEA* allele [8,9] in the central cell of the female gametophyte. Silencing of the paternal *MEA* allele is not directly controlled by DNA methylation. Instead, PcG proteins, including maternally expressed MEA, repress expression of the paternal *MEA* allele in the endosperm [9]. Thus, *MEA* is a self-imprinted gene [9–11]. DME demethylation occurs by the base excision DNA repair process [9,12]. DME first excises 5-methylcytosine by its glycosylase activity, and then the AP lyase activity of DME nicks the DNA strand. AP endonuclease cleavage generates a 3' hydroxyl, DNA polymerase replaces the excised 5-methylcytosine with cytosine, and finally DNA ligase seals the nick. By excising 5-methylcytosine, DME prevents CpG hypermethylation of its target genes, thus activating their gene expression. DME is specifically expressed in the central cell

of the female gametophyte before fertilization [8]. Its temporal and spatial expression is essential for the establishment of *MEA* imprinting and seed viability.

In mammals, there are approximately 100 imprinted genes with approximately half expressing the maternal allele and half expressing the paternal allele (http://www.geneimprint.com). In *Arabidopsis*, only four genes imprinted in the endosperm have been discovered [7,13–15]. *MEA*, *FWA*, and *FIS2* are expressed maternally and are controlled by DME [8,9,13,15]. *PHERES1 (PHE1)* is expressed paternally and PcG proteins including MEA, which is activated by DME, regulate its maternal silencing [14,16]. Therefore, DME is a key regulator of genomic imprinting in flowering plants.

Considering the number of imprinted genes in animals, it is possible that there are more imprinted genes in *Arabidopsis* that are controlled by DME. To test this hypothesis, we used an RNA profiling approach to search for other genes regulated by DME. We generated transgenic plants that ectopically express DME under the control of the *CaMV* promoter in pollen and stamen [17]. Since the endogenous *DME* gene is not expressed in the wild type male reproductive organs (pollen and stamen), ectopic DME expression reveals the expression of novel target genes. We extracted poly(A)+ RNA from pollen and stamens harvested from *CaMV:DME* transgenic plants and wild type control plants. cRNAs were labeled and hybridized to Affymetrix *Arabidopsis* ATH1 GeneChip arrays. We compared and analyzed RNA profiles from both transgenic and wild type pollen and stamen and detected candidate DME-inducible genes, which were confirmed by semiquantitative RT-PCR experiments. These DME-inducible genes shed light on the mechanism of DME-mediated demethylation and the role of gene imprinting in seed biology.

Materials and methods

Plant materials, transgenic plant isolation, and growth condition. *CaMV:DME*-4 transgenic plants were generated as described [17] and wild type plants (Columbia-*gl* ecotype) were used in this experiment. Transgenic plants contained the *NPTII* gene, which confers resistance to kanamycin. Plants were grown in standard greenhouse conditions (16 h light/8 h dark).

Pollen collection and RNA extraction. Pollen grains and stamens were collected from stage 13 and 14 flowers and processed as described [17,18]. Total RNAs were extracted using TRIzol (Invitrogen) as described [7]. Poly(A)+ RNA was selected from total RNA by using the Dynabeads mRNA Purification kit (Dynal Inc.) according to procedures provided by the manufacturer.

Preparation of the probes, hybridization, and scanning. Double-stranded cDNAs were generated using Superscript Choice system (Invitrogen) and were purified by phenol/chloroform extraction. Biotin-labeled cRNA probes were synthesized using BioArray High Yield RNA Transcript Labeling Kit (T7) (Enzo Life Sciences) and purified using RNeasy mini spin columns (Qiagen). Concentration of cRNA probe was determined using a UV-spectrophotometer and the size range of synthesized cRNA was determined by fractionating cRNAs on a 1.3% agarose/formaldehyde gel. cRNAs were fragmented with fragmentation buffer (40 mM Tris–acetate, pH 8.1, 100 mM KOAc, and 30 mM MgOAc) and cRNA size was

determined on a 1.3% agarose/formaldehyde gel. Fragmented biotin-labeled cRNAs were hybridized to the *Arabidopsis* ATH1 GeneChip Array (Affymetrix) for 16 h according to manufacturer's protocol (Affymetrix). The arrays were stained with streptavidin phycoerythrin (SAPE, Molecular probes), goat IgG (Sigma), and anti-streptavidin biotinylated antibody. GeneChip arrays were scanned and analyzed using Affymetrix Microarray Analysis Suite (MAS) 5.0 (Santa Clara, CA). All probe sets on the array were scaled globally to a target intensity of 500 using MAS 5.0 default parameters. All GeneChip data reported in this paper have been deposited in the GEO database (http://www.ncbi.nlm.nih.gov/geo) as series GSE9408.

RT-PCR analysis. Reverse transcription was done with Retro Transcript (Ambion) following instructions provided by the manufacturer. Primers used in these experiments are in Supplementary Materials.

Results and discussion

DME expression in CaMV:DME pollen

We previously generated transgenic plants bearing a transgene (*CaMV:DME*) with the 6.8 kb full-length *DME* cDNA ligated to the cauliflower mosaic virus promoter (CaMV) [17]. In multiple *CaMV:DME* transgenic lines, DME was ectopically expressed in stamen and pollen, and we detected expression of known DME target genes, *MEA* [17] and *FWA* (data not shown). This suggests that DME expression is both necessary and sufficient to activate target gene expression, and ectopic DME expression in stamen or pollen can induce DME target genes that are normally silenced in the male reproductive organs [17].

Transcriptional profiling of CaMV:DME and wild type

To gain a global view of the transcription network that is regulated by DME, we used Affymetrix GeneChip arrays to identify novel genes up-regulated by ectopic DME expression in *CaMV:DME* pollen and stamen compared to the wild type counterparts. Pollen grains were harvested from *CaMV:DME* and control wild type flowers at stages 13 (open flowers) and 14 (self-pollinating) [19]. We also collected stamens at the same stages from wild type and *CaMV:DME* plants. Stamens included filaments, and anthers containing mature and immature pollen grains. Poly(A)+ RNAs were extracted from the above four independent tissue samples, and their biotin-labeled cRNA probes were synthesized from double-stranded cDNAs generated from poly(A)+ RNAs. Biotinylated cRNA probes were hybridized to Affymetrix *Arabidopsis* ATH1 GeneChip arrays and data were analyzed using the GeneSpring program. Scatter plot analysis shows that wild type and *CaMV:DME* RNA profiles from pollen are well correlated ($r = 0.99$) (Fig. 1A). Similar results were seen when wild type and *CaMV:DME* stamen RNAs were compared ($r = 0.95$) (Fig. 1B). The high correlation coefficients between wild type and *CaMV:DME* pollen and stamen indicate that ectopic expression of DME induces a small number of genes.

From the transcriptional profiling data, 6392 and 7281 RNAs were detected in wild type and *CaMV:DME* pollens, respectively (Fig. 2A). These numbers are in close
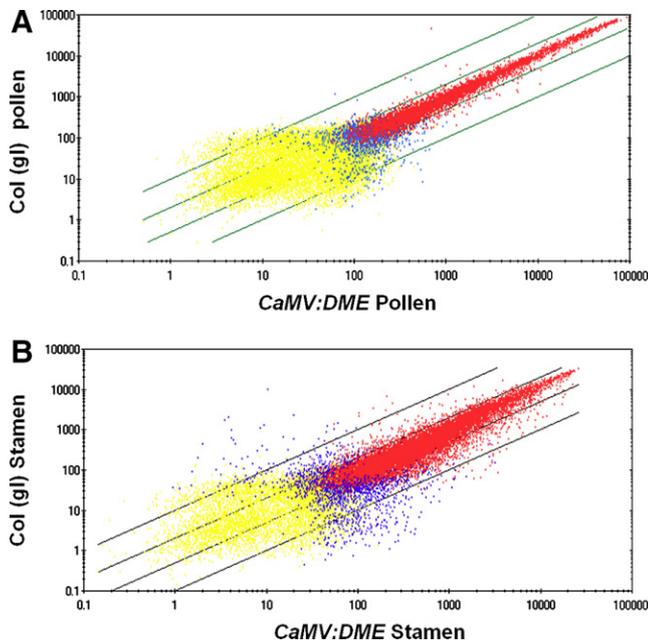
Fig. 1. Scatter plots of wild type and *CaMV*:DME pollen and stamen. Scatter plot analysis of wild type and *CaMV:DME pollen (A) and stamen (B)* was generated using MAS 5.0 of all probe sets represented on the Affymetrix *Arabidopsis* ATH1 GeneChip array. A red dot indicates RNA with a detection call of present by MAS 5.0 in both wild type and *CaMV:DME* pollen or stamen. A yellow dot indicates RNA with a detection call of absent in both wild type and *CaMV:DME* pollen or stamen. Blue dots indicate different detection calls by MAS 5.0 for wild type and *CaMV:DME* pollen and stamen. (For interpretation of the references in color in this figure legend, the reader is referred to the web version of this article.)
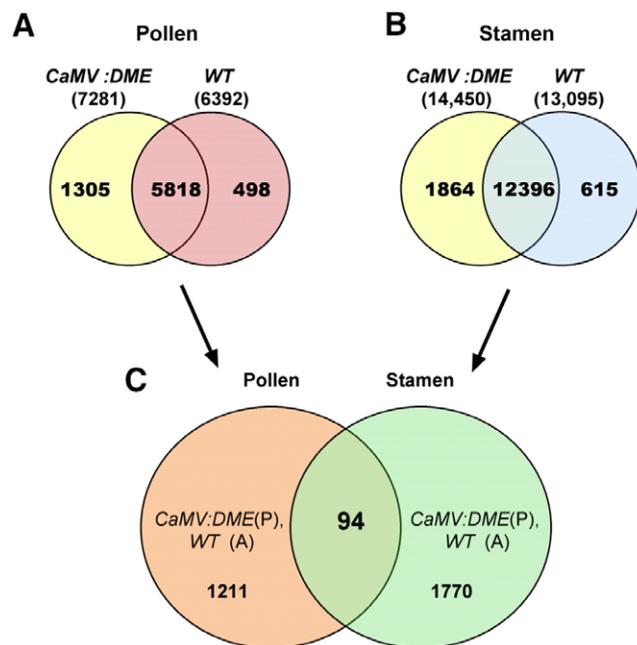

Fig. 2. Analysis of RNA profiles. Venn diagrams of genes expressed in wild type and *CaMV:DME* pollen (A), and wild type and *CaMV:DME* stamen (B). Ninety-four genes show DME-dependent expression in both *CaMV:DME* stamen and pollen (C).

agreement with RNA profiling data of pollen obtained by others [20,21]. RNAs (1305) were detected in *CaMV:DME* pollen and not in wild type control pollen. Similarly, 13,095 and 14,450 RNAs were detected in wild type and *CaMV:DME* stamens, respectively (Fig. 2B). These numbers are in close agreement with RNA profiling data from stamens obtained by others [22]. RNAs (1864) were detected in *CaMV:DME* but not in wild type stamens. We found 94 genes with DME-inducible expression in both *CaMV:DME* stamen and pollen when comparing *CaMV:DME* pollen- and stamen-specific RNAs (Fig. 2C). Thus, only a small number of DME-inducible genes were detected in both pollen and stamen.

These 94 genes were sorted into different functional categories (Supplementary Table 1). There are 4 genes related to cell growth and division, 5 genes to cell structure, 3 genes to disease and defense, 3 genes for energy, 2 genes for intracellular traffic, 9 genes for metabolism, 4 genes for post-transcription, 7 genes for protein destination and storage, 1 gene for protein synthesis, 3 genes for secondary metabolism, 10 genes for signal transduction, 14 genes for transcription, 4 genes for transporter, 2 genes for transposon, and 23 genes unclassified (Supplementary Table 1).

We did not detect expression of *MEA* or *FWA* in the RNAs isolated from *CaMV:DME* pollen or stamen. It is possible that the level of *MEA* and *FWA* RNA was too low to be detected by the *Arabidopsis* whole genome Affymetrix probe array [23].

We also sorted the 94 genes into gene ontologies using the bulk retrieval tool at TAIR (http://www.arabidopsis.org/tools/bulk/go) (Fig. 3). The overall transcription profile is similar to the whole *Arabidopsis* genome profile [24] suggesting DME might regulate genes with a broad range of functional categories. Gene ontology analysis identified three molecular functions enriched in the DME-inducible RNA list compared to the whole *Arabidopsis* genome: DNA or RNA binding (11.1% versus 5.4%), kinase activity (8.3% versus 5.0%), and transcription factor activity (8.3% versus 5.0%). Although DME induces a wide range of genes, this enrichment suggests that target genes may share common regulatory pathways or be involved in similar functions such as DNA or RNA binding, kinase activity, and transcription factor activity.

In *Arabidopsis*, a number of genes are demethylated by ROS1, DML2, and DML3 [25,26]. ROS1 (REPRESSOR OF SILENCING 1) is a DME homolog that maintains the transcriptionally active states of a *RD29A::LUCIFERASE (RD29A::LUC)* reporter gene and the endogenous *RD29A* gene [12,27,28]. DML2 (DEMETER-LIKE 2) and DML3 (DEMETER-LIKE 3) are two other DME-like genes in *Arabidopsis*. Genes regulated by ROS1 have been identified by comparing RNA profiles from wild type and *ros1* mutant plants [26]. Genes demethylated by ROS1, DML2, and DML3 have been revealed by comparing wild type and *ros1*, *dml2*, and *dml3* triple mutant genome-tiling microarrays [25]. We found little overlap between our DME target genes and these published results. Thus,
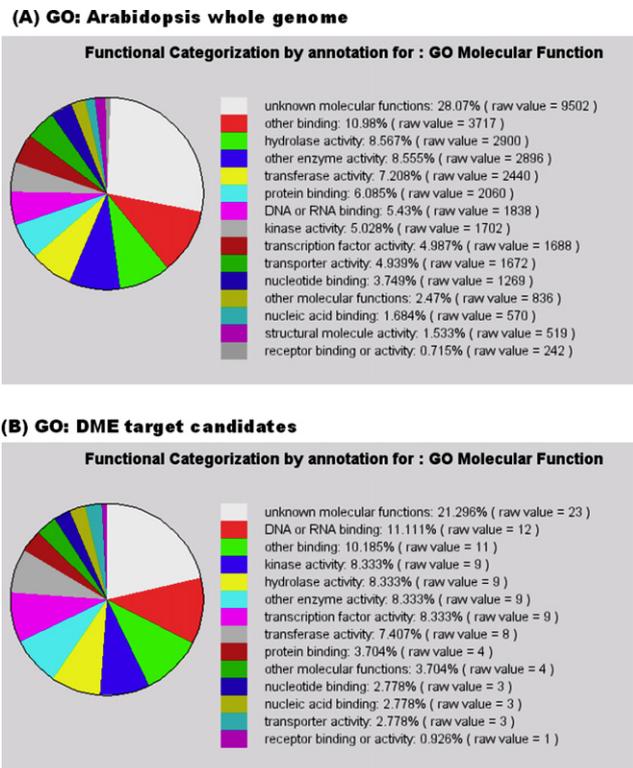
Fig. 3. Functional categorization of wild type *Arabidopsis* whole genome (A) and genes induced by DME (B). Ninety-four genes were sorted out to the functional categories classified in the databases of gene ontology at TAIR. DME-dependent downstream genes show similar distribution of molecular function with the *Arabidopsis* whole genome.
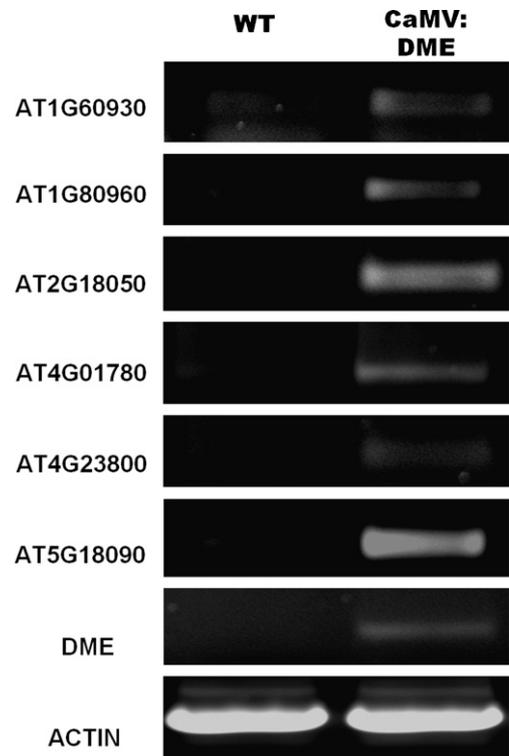

Fig. 4. Semi-quantitative RT-PCR validation of putative DME target genes. Expression patterns of the possible DME-downstream genes were confirmed by semi-quantitative RT-PCR from independently isolated pollen tissue.

although DME and DMLs have similar DNA demethylase activity [9,12,25], their putative target genes are very distinctive based on the data from microarray and genome-tiling analyses.

Semi-quantitative RT-PCR for the possible candidate genes

We chose six of the 94 genes that represent a wide range of biological processes for independent validation. Semi-quantitative RT-PCR was performed to measure the RNA level of putative DME-downstream genes from wild type and *CaMV:DME* pollen tissue. We found that RT-PCR results of these six genes were consistent with the microarray gene profiling results (Fig. 4). That is, expression was not detected in control wild type pollen and was induced in *CaMV:DME* pollen. The list of representative candidate DME target genes includes: *AT1G60930* which encodes a DEAD/DEAH box domain RecQ family DNA helicase involved in DNA replication, recombination, RNA metabolism, and gene silencing [29]; *AT1G80960* which encodes a F-box-related protein with leucine-rich repeat domain; *AT2G18050* that encodes a linker Histone H1-3, involved in DNA binding and chromatin structure that is induced by dehydration and ABA [30]; *AT4G01780,* encoding a XH/XS domain-containing protein; *AT4G23800,* a high mobility group (HMG) family

protein and a non-histone component of chromatin that is involved in DNA binding inducing DNA-dependent transcription, replication, and repair mechanism [31]; and *AT5G18090,* encoding a transcription factor B3 family protein. Taken together, these results suggest that DME, either directly or indirectly, regulates expression of the above genes in addition to *MEA*, *FWA*, and *FIS2*. Since their induction in expression is in response to ectopic expression of DME in tissues where DME is not normally active, they might be bona fide target genes of DME.

Whether DME changes the methylated status of each downstream gene, their biological function, and whether they are imprinted genes remain to be assessed. These future studies will provide a more global picture of gene imprinting and will help us to understand the mechanism whereby DME activates gene expression by DNA demethylation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2007.10.092.

References

[1] A.L. Fowden, C. Sibley, W. Reik, M. Constancia, Imprinted genes, placental development and fetal growth, Horm. Res. 65 (2006) 50–58.

[2] M. Gehring, Y. Choi, R.L. Fischer, Imprinting and seed development, Plant Cell 16 (2004), doi:10.1105/tpc.017988.

[3] R.J. Scott, M. Spielman, Genomic imprinting in plants and mammals: how life history constrains convergence, Cytogenet. Genome Res. 113 (2006) 53–67.

[4] U. Grossniklaus, J.P. Vielle-Calzada, M.A. Hoeppner, W.B. Gagliano, Maternal control of embryogenesis by *MEDEA*, a polycomb group gene in *Arabidopsis*, Science 280 (1998) 446–450.

[5] T. Kiyosue, N. Ohad, R. Yadegari, M. Hannon, J. Dinneny, D. Wells, A. Katz, L. Margossian, J. Harada, R.B. Goldberg, R.L. Fischer, Control of fertilization-independent endosperm development by the *MEDEA* polycomb gene in *Arabidopsis*, Proc. Natl. Acad. Sci. USA 96 (1999) 4186–4191.

[6] M. Luo, P. Bilodeau, A. Koltunow, E.S. Dennis, W.J. Peacock, A.M. Chaudhury, Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*, Proc. Natl. Acad. Sci. USA 96 (1999) 296–301.

[7] T. Kinoshita, R. Yadegari, J.J. Harada, R.B. Goldberg, R.L. Fischer, Imprinting of the *MEDEA* polycomb gene in the *Arabidopsis* endosperm, Plant Cell 11 (1999) 1945–1952.

[8] Y. Choi, M. Gehring, L. Johnson, M. Hannon, J.J. Harada, R.B. Goldberg, S.E. Jacobsen, R.L. Fischer, DEMETER, a DNA Glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*, Cell 110 (2002) 33–42.

[9] M. Gehring, J.H. Huh, T.F. Hsieh, J. Penterman, Y. Choi, J.J. Harada, R.B. Goldberg, R.L. Fischer, DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation, Cell 124 (2006) 495–506.

[10] C. Baroux, V. Gagliardini, D.R. Page, U. Grossniklaus, Dynamic regulatory interactions of Polycomb group genes: MEDEA autoregulation is required for imprinted gene expression in Arabidopsis, Genes Dev. 20 (2006) 1081–1086.

[11] P.E. Jullien, A. Katz, M. Oliva, N. Ohad, F. Berger, Polycomb group complexes self-regulate imprinting of the Polycomb group gene *MEDEA* in *Arabidopsis*, Curr. Biol. 16 (2006) 486–492.

[12] T. Morales-Ruiz, A.P. Ortega-Galisteo, M.I. Ponferrada-Marin, R.R. Martinez-Macias, R.R. Ariza, T. Roldan-Arjona, *DEMETER* and *REPRESSOR OF SILENCING1* encode 5-methylcytosine DNA glycosylases, Proc. Natl. Acad. Sci. USA 103 (2006) 6853–6858.

[13] T. Kinoshita, A. Miura, Y. Choi, Y. Kinoshita, X. Cao, S.E. Jacobsen, R.L. Fischer, T. Kakutani, One-way control of *FWA* imprinting in *Arabidopsis* endosperm by DNA methylation, Science 303 (2004) 521–523.

[14] C. Kohler, D.R. Page, V. Gagliardini, U. Grossniklaus, The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of *PHERES1* by parental imprinting, Nat. Genet. 37 (2005) 28–30.

[15] P.E. Jullien, T. Kinoshita, N. Ohad, F. Berger, Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting, Plant Cell 18 (2006) 1360–1372.

[16] G. Makarevich, O. Leroy, U. Akinci, D. Schubert, O. Clarenz, J. Goodrich, U. Grossniklaus, C. Kohler, Different Polycomb group complexes regulate common target genes in *Arabidopsis*, EMBO Rep. 7 (2006) 947–952.

[17] Y. Choi, J.J. Harada, R.B. Goldberg, R.L. Fischer, An invariant aspartic acid in the DNA glycosylase domain of DEMETER is necessary for transcriptional activation of the imprinted *MEDEA* gene, Proc. Natl. Acad. Sci. USA 101 (2004) 7481–7486.

[18] D. Preuss, B. Lemieux, G. Yen, R.W. Davis, A conditional sterile mutation eliminates surface components from *Arabidopsis* pollen and disrupts cell signaling during fertilization, Genes Dev. 7 (1993) 974–985.

[19] J.L. Bowman, S.G. Mansfield, Embryogenesis, in: J. Bowman (Ed.), *Arabidopsis*: An Atlas of Morphology and Development, Springer-Verlag, New York, 1994, pp. 351–361.

[20] D. Honys, D. Twell, Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*, Genome Biol. 5 (2004) R85.

[21] C. Pina, F. Pinto, J.A. Feijo, J.D. Becker, Gene family analysis of the *Arabidopsis* pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation, Plant Physiol. 138 (2005) 744–756.

[22] A. Mandaokar, B. Thines, S. Byongchul, B.M. Lange, G. Choi, Y.J. Koo, Y.J. Yoo, Y.D. Choi, G. Choi, J. Browse, Transcriptional regulators of stamen development in *Arabidopsis* identified by transcriptional profiling, Plant J. 46 (2006) 984–1008.

[23] J.C. Redman, B.J. Haas, G. Tanimoto, C.D. Town, Development and evaluation of an *Arabidopsis* whole genome Affymetrix probe array, Plant J. 38 (2004) 545–561.

[24] T.Z. Berardini, S. Mundodi, R. Reiser, E. Huala, M. Garcia-Hernandez, P. Zhang, L.M. Mueller, J. Yoon, A. Doyle, G. Lander, N. Moseyko, D. Yoo, I. Xu, B. Zoeckler, M. Montoya, N. Miller, D. Weems, S.Y. Rhee, Functional annotation of the *Arabidopsis* genome using controlled vocabularies, Plant Physiol. 135 (2004) 1–11.

[25] J. Penterman, D. Zilberman, J.H. Huh, T. Ballinger, S. Henikoff, R.L. Fischer, DNA demethylation in the *Arabidopsis* genome, Proc. Natl. Acad. Sci. USA 104 (2007) 6752–6757.

[26] J. Zhu, A. Kapoor, V.V. Sridhar, F. Agius, J.K. Zhu, The DNA glycosylase/lyase ROS1 functions in pruning DNA methylation patterns in *Arabidopsis*, Curr. Biol. 17 (2007) 54–59.

[27] Z. Gong, T. Morales-Ruiz, R.R. Ariza, T. Roldan-Arjona, L. David, J.J. Zhu, *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA Glycosylase/Lyase, Cell 111 (2002) 803–814.

[28] F. Agius, A. Kapoor, J.K. Zhu, Role of the *Arabidopsis* DNA glycosylase/lyase ROS1 in active DNA demethylation, Proc. Natl. Acad. Sci. USA 103 (2006) 11796–11801.

[29] F. Hartung, H. Plchova, H. Puchta, Molecular characterisation of RecQ homologues in *Arabidopsis thaliana*, Nucleic Acid Res. 28 (2000) 4275–4282.

[30] R. Ascenzi, J.S. Gantt, Molecular genetic analysis of the drought-inducible linker histone variant in *Arabiidopsis thaliana*, Plant Mol. Biol. 41 (1999) 159–169.

[31] J.L. Riechmann, J. Heard, G. Martin, L. Reuber, C. Jiang, J. Keddie, L. Adam, O. Pineda, O.J. Ratcliffe, R.R. Samaha, R. Creelman, M. Pilgrim, P. Broun, J.Z. Zhang, D. Ghandehari, B.K. Sherman, G. Yu, *Arabidopsis* transcription factors: genome-wide comparative analysis among eukaryotes, Science 290 (2000) 2105–2110.