Studies on DNA Methylation and Demethylation-mediated Epigenetic Genome Regulation in Rice

（イネにおけるDNAメチル化および脱メチル化を介したエピジェネティックなゲノム制御機構の研究）

AKEMI ONO

Doctoral Program in Life and Food Sciences

Graduate School of Science and Technology

Niigata University

Contents

Abbreviations

bp	base pair
BER	base excision repair
CentO	Centromeric tandem repeats O
CMT3	CHROMOMETHYLASE 3
DAP	day after pollination
DAPI	4', 6-diamidino-2-phenylindole
Dart	*DNA-based active rice transposon*
DDM1	DECREASE IN DNA METHYLATION 1
DME	DEMETER
DML2	DEMETER-LIKE 2
DML3	DEMETER-LIKE 3
DRM2	DOMAINS REARRANGED METHYLTRANSFERASE 2
DT-A	*diphtheria toxin A fragment*
Δ*En*	the 3' end of maize *Enhancer* element
FIE1	FERTILIZATION INDEPENDENT ENDOSPERM 1
FIS2	FERTILIZATION INDEPENDENT SEED 2
GFP	green fluorescent protein
GT	gene targeting
GUS	β-glucuronidase
hpt	hygromycin phosphotransferase
HR	homologous recombination
IPTG	isopropyl β-D-1-thiogalactopyranoside
JF	junction fragment
LB	left border
Lsh	Lymphocyte-specific helicase
LTR	long terminal repeat
MEA	MEDEA
5-meC	5-methyl cytosine
MET1	DNA METHYLTRANSFERASE 1
MuDR	*Mutator* transposon regulatory element
mPing	*miniture Ping* transposable element

PRC2	Polycomb Repressive Complex 2
RAP database	Rice Annotation Project database
RB	right border
rDNA	ribosomal DNA
RIRE7	*RIce REtrotransposon 7*
RNAi	RNA interference
ROS1	REPRESSOR OF SILENCING 1
RT-PCR	reverse transcription-PCR
SC	sperm cell
TE	transposable element
TOS17	retrotransposon *TOS17*
VN	vegetative nucleus
WT	wild type

General Introduction

The genetic information must be transmitted accurately through cell divisions (mitosis) and generations (meiosis). The genetic information in a cell is stored in DNA, which is packaged around histones into chromatin (Figure 1a). Epigenetic modifications of DNA and histones, which are stable and heritable chemical modifications, influence the expression of the underlying genes without altering its nucleotide structure and constitute an additional layer of information in genome function (Figure 1b). One such epigenetic modification is DNA methylation, which is the addition of a methyl group to DNA residues, and occurs primarily at the carbon-5 position of cytosines in eukaryotes. Cytosine methylation is evolutionarily ancient and, apart from yeast and some invertebrates, present at detectable levels in all organisms that have been examined to date. In addition, the defects of cytosine methylation in mammals lead to embryonic lethal and in plants they can cause pleiotropic morphological defects, which attest to the important roles of this modification. Cytosine methylation, which is the most frequent in a CG sequence context in plant and animal genomes and also found in the CHG (H = A, C, or T) and CHH contexts in plants, are established and maintained by several components (reviewed in Law and Jacobsen, 2010, Table 1). In plants, genes that mediate these processes have been characterized mainly

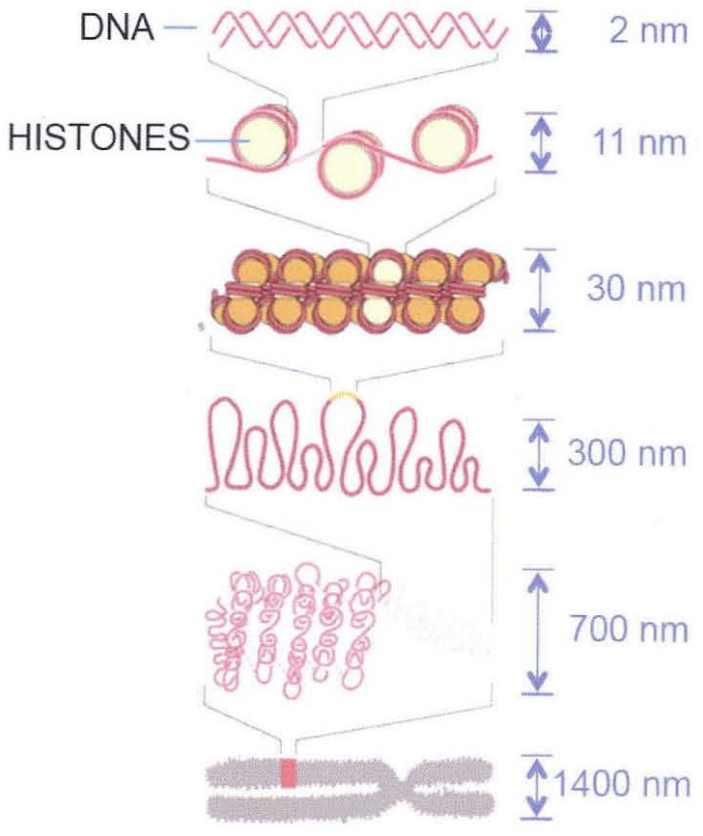
in Arabidopsis (*Arabidopsis thaliana*). *De novo* methylation is catalyzed by DOMAINS REARRANGED METHYLTRANSFERASE 2 (DRM2), and maintained by three different pathways; CG methylation is maintained by DNA METHYLTRANSFERASE 1 (MET1), CHG methylation is maintained by CHROMOMETHYLASE 3 (CMT3), and CHH methylation is maintained through persistent *de novo* methylation by DRM2 (reviewd in Henderson and Jacobsen, 2007). DECREASE IN DNA METHYLATION 1 (DDM1), a SWI2/SNF2 chromatin remodeling factor, is also required for the maintenance of cytosine methylation by unknown mechanisms (Vongs *et al.*, 1993; Jeddeloh *et al.*, 1999; Brzeski and Jerzmanowski, 2003). DNA methylation can be also removed either passively, when the maintenance methylation that usually follow a DNA replication is blocked, or by a more active process when methyl cytosine is enzymatically removed. Active demethyaltion is achieved by bifunctional DNA glycosylase activity in plants. The Arabidopsis glycosylases, DEMETER (DME), REPRESSOR OF SILENCING 1 (ROS1), DEMETER-LIKE 2 (DML2) and DML3, recognize and remove methylated cytosines with base excision repair (BER) pathway and replace into cytosines with DNA polymerase and DNA ligase activity (Gehring *et al.*, 2009; Zhu, 2009, Figure 2).

In spite of the importance as fundamental process, the study of DNA methylation in plants is primarily limited in Arabidopsis. In this thesis, I addressed

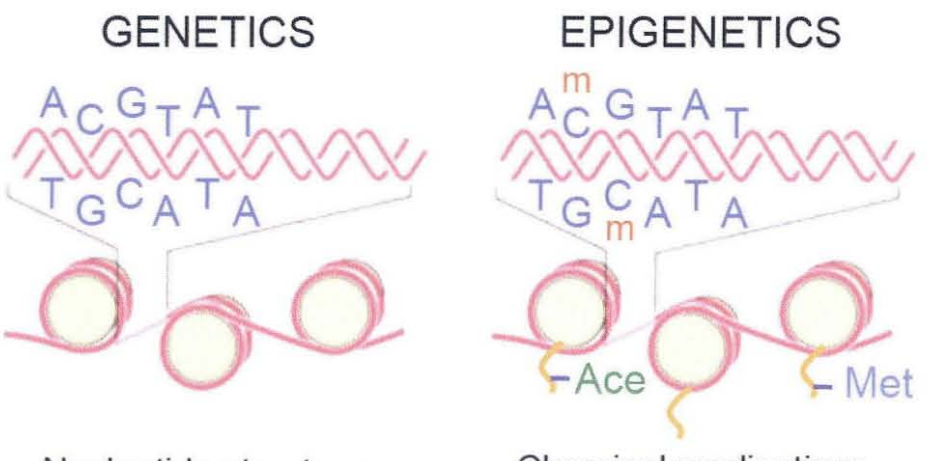
the questions in order to understand the genome regulation mediated by DNA methylation/demethylation in rice (*Oryza sativa*), which is a model cereal for agribioscience as well as a major crop for resource of calories. As an well established model system for crop, rice has several advantages to study molecular mechanisms; completed genome sequence, accessible databases, easy transformation system, and applicable homologous recombination (HR)-promoted gene targeting system (Terada *et al.*, 2002, 2007; Johzuka-Hisatomi *et al.*, 2008; Yamauchi *et al.*, 2009), which has been providing powerful tools as routine practice in mice to understand the molecular mechanisms behind epigenetics. By employing a reproducible gene targeting procedure, which can modify any endogenous genes as we designed, in principle, transgenic rice carrying an identical and desired gene allele can be generated. Taking these advantages of rice as a model crop, I planned to study DNA methylation through reverse genetics. In Chapter 1, in order to understand the impact of cytosine demethylation on rice genome, knock-in plants, which have disrupted *ROS1a* gene fused with *GUS* marker gene to monitor the knock-in allele, were generated and analyzed in detail. In Chapter 2, *DDM1a* and *DDM1b* knock-out plants were generated. Analyzing *DDM1a*, *DDM1b* and *DDM1a/DDM1b* double disruptant mutants, I addressed the role of cytosine methylation in rice genome. Through these studies, I discuss about DNA

methylation/demethylation-mediated epigenetic genome regulation in rice. I believe these studies would provide valuable information not only for basic science of rice, but also for applied crop research.

(a)



(b)



Nucleotide structure of DNA

Chemical modications of DNA and histones

Figure 1

Figure 1. The genetic and epigenetic information in a cell.

(a) DNA, which encodes genetic information in a cell, is packaged into chromatin. Chromatin is packaged further in highly order in a cell.

(b) The genetic information is encoded by primary DNA sequence (left). Chemical modifications of DNA and histones, the core components of chromatin, constitute epigenetic information (right). m, addition of a methyl group to a cytosine base; Ace, addition of an acetyl group to histones; Met, addition of a methyl group to histones.

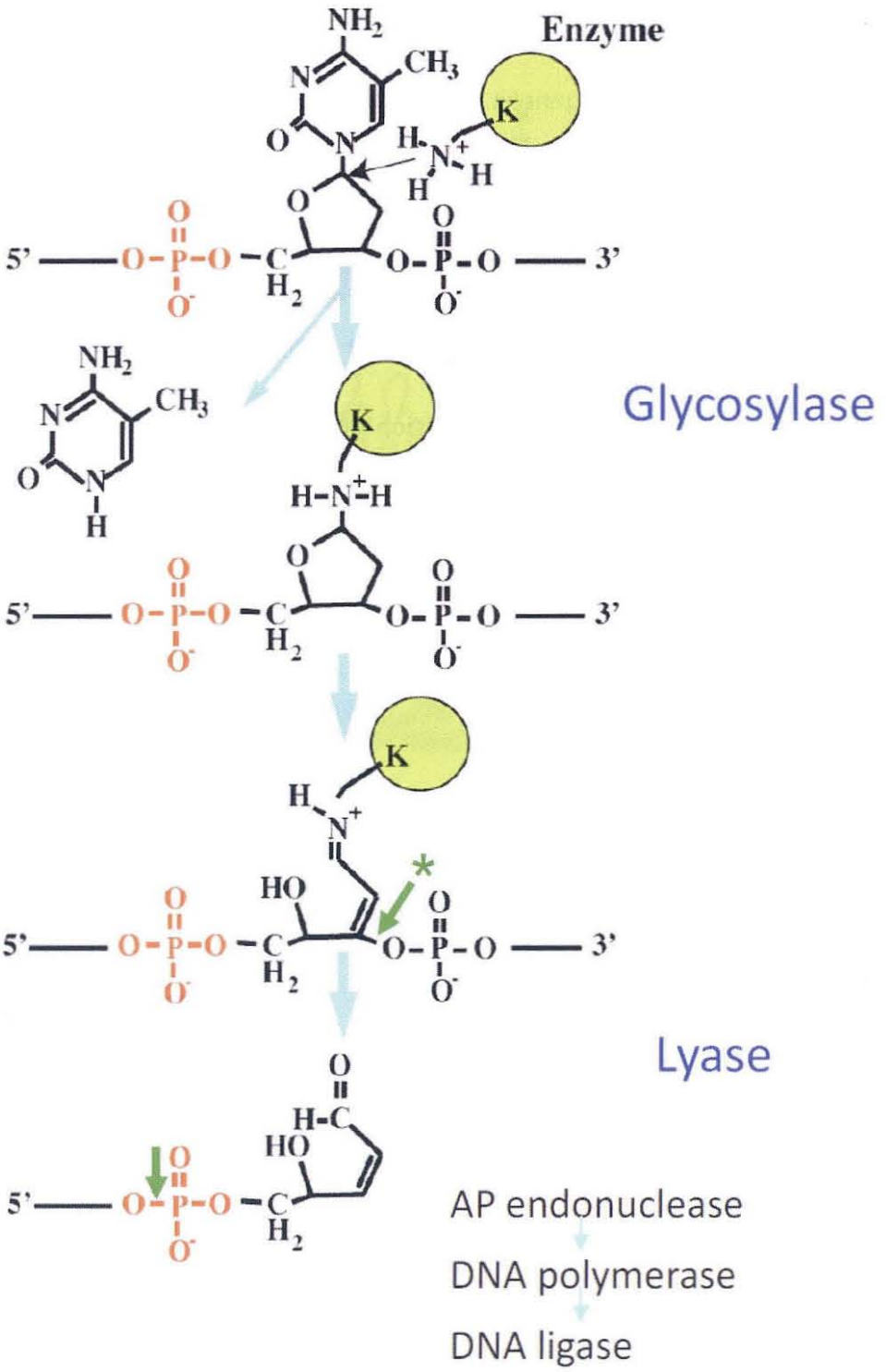

Enzyme
Glycosylase
Lyase
AP endonuclease
DNA polymerase
DNA ligase

Figure 2

Figure 2. Schematic mechanism of bisfunctional DNA glycosylases.

A conserved aspartic acid of the glycosylase (green circle) acquires a proton from a conserved lysine residue (K in green circle) that attacks the C1' carbon of the deoxyribose ring, creating a covalent DNA-enzyme intermediate. β or σ elimination reactions release the enzyme from the DNA and cleave one of the phosphodiester bonds (green arrow with asterisk). Cleavage 5' to the abasic site of the β or σ elimination (green arrow) produced by an AP endonuclease generates a 3'-hydroxyl group used by a DNA repair polymerase that insert the proper nucleotide, and a DNA ligase seals the nick. Modified from Gehring *et al.* (2006).

Table 1. Proteins involved in *de novo* DNA methylation, maintenance methylation and demethylation. Selected proteins with established roles in these processes are shown for mice [(*Mus musculum* (*Mm*)], *Arabidopsis thaliana* (*At*), rice [*Oryza sativa* (*Os*)], and zebrafish [*Danio rerio* (*Dr*)]. Modified from Law and Jacobsen (2010).

Category	Name	Function
De novo methylation	*Mm*DNMT3L	Non-catalysic factor
	*Mm*DNMT3A *Mm*DNMT3B *At*DRM2 *Os*DRM2*	*De novo* DNA methyltransferases
	*Mm*MIWI2 *Mm*MILI *At*AGO	Argonaute family small interfering RNA-binding proteins
Maintenance methylation	*Mm*DNMT1 *At*MET1	Maintenance DNA methyltransferases
	*Mm*UHRF1 *At*VIM1, *At*VIM2, *At*VIM3	Chromatin-binding proteins proposed to recruit maintenance DNA methyltransferases
	*Mm*LSH1 *At*DDM1	Chromatin-remodeling factors
	*At*CMT3	Plant-specific DNA methyaltransferase
	*At*SUVH4, *At*SUVH5, *At*SUVH6	Histone methyltransferases (histone 3 lysine 9) with SRA methyal-DNA-binding domains
Demethylation	*At*DME *At*ROS1 *At*DML2 *At*DML3 DNG701** (= LOC_Os05g37350)	5-methylcytosine glycosylases
	*Mm*MBD4 *Mm*TDG	Thymine DNA glycosylases
	*Mm*AID *Mm*APOBEC *Dr*Aid *Dr*Apobec2	Cytosine and 5-methylcytosine deaminases

* Moritoh *et al.*, 2012, ** La *et al.*, 2011.

REFERENCES

Brzeski, J. and Jerzmanowski, A. (2003) *Deficient in DNA Methylation 1 (DDM1)* defines a novel family of chromatin-remodeling factors. *J. Biol. Chem.* **278**, 823-828.

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

Gehring, M., Reik, W. and Henikoff, S. (2009) DNA demethylation by DNA repair. *Trends Genet.* **25**, 82-90.

Henderson, I.R. and Jacobsen, S.E. (2007) Epigenetic inheritance in plants. *Nature*, **447**, 418-424.

Hsieh, T.-F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L. and Zilberman, D. (2009) Genome-wide demethylation of *Arabidopsis* endosperm. *Science*, **324**, 1451-1454.

Jeddeloh, J.A., Stokes, T.L. and Richards, E.J. (1999) Maintenance of genomic methylation requires a SWI2/SNF2-loke protein. *Nat. Genet.* **22**, 94-97.

Johzuka-Hisatomi, Y., Terada, R. and Iida, S. (2008) Efficient transfer of base changes from a vector to the rice genome by homologous recombination:

involvement of heteroduplex formation and mismatch correction. *Nucleic Acids Res.* **36**, 4727-4735.

La, H., Ding, B., Mishra, G.P., *et al.* (2011) A 5-methylacytosine DNA glycosylase/lyase demethylates the retrotransposon *TOS17* and promotes its transposition in rice. *Proc. Natl. Acad. Sci. USA*, **108**, 15498-15503.

Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204-220.

Moritoh, S., Eun, C-H., Ono, A., Asao, H., Okano, Y., Yamaguchi, K., Shimatani, Z., Koizumi, A. and Terada, R. (2012) Targeted disruption of orthologue of *DOMAINS REARRANGED METHYLASE 2*, *OsDRM2*, impairs the growth of rice plants by abnormal DNA methylation. *Plant J.* **71**, 85-98.

Terada, R., Urawa, H., Inagaki, Y., Tsugane, K. and Iida, S. (2002) Efficient gene targeting by homologous recombination in rice. *Nat. Biotechnol.* **20**, 1030-1034.

Terada, R., Johzuka-Hisatomi, Y., Saitoh, M., Asao, H. and Iida, S. (2007) Gene targeting by homologous recombination as a biotechnological tool for rice functional genomics. *Plant Physiol.* **144**, 846-856.

Vongs, A., Kakutani, T., Martienssen R.A. and Richards, E.J. (1993) *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**, 1926-1928.

Yamauchi, T., Johzuka-Hisatomi, Y., Fukada-Tanaka, S., Terada, R., Nakamura, I. and Iida, S. (2009) Homologous recombination-mediated knock-in targeting of the *MET1a* gene for a maintenance DNA methyltransferase reproducibly reveals dosage-dependent spatiotemporal gene expression in rice. *Plant J.* **60**, 386-396.

Zhu, J.-K. (2009) Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43,** 143-166.

Chapter 1

A null mutation of *ROS1a* for DNA demethylation in rice is not transmittable to progeny

INTRODUCTION

Cytosine DNA methylation, which is found in the CG, CHG (H = A, C, or T), and CHH sequence contexts, is an epigenetic modification in plants (Martienssen and Colot, 2001; Henderson and Jacobsen, 2007). The genes and their corresponding encoded enzymes that mediate DNA methylation and demethylation have been characterized mainly in Arabidopsis (Chan *et al.*, 2005; Zhu, 2009; Law and Jacobsen, 2010). The Arabidopsis enzymes that mediate 5-methylcytosine (5-meC) DNA demethylation [DEMETER (DME) (Choi *et al.*, 2002; Gehring *et al.*, 2006), REPRESSOR OF SILENCING 1 (ROS1) (Gong *et al.*, 2002; Agius *et al.*, 2006), DEMETER-LIKE 2 (DML2), and DML3 (Choi *et al.*, 2002; Ortega-Galisteo *et al.*, 2008)] are bifunctional DNA glycosylases that not only recognize and remove 5-meC from double-stranded DNA but also carry lyase activity that nicks the double-stranded DNA at the abasic site (Gehring *et al.*, 2009b; Zhu, 2009; Law and Jacobsen, 2010).

The largest gene among these DNA glycosylases, *DME*, is predominantly expressed in the homodiploid central cell of the female gametophyte before fertilization. DME promotes maternal allele–specific global hypomethylation and the expression of both imprinted genes and elements including transposons in the endosperm (Choi *et al.*, 2002; Gehring *et al.*, 2006, 2009a). In contrast, *ROS1, DML2*, and *DML3* are expressed in vegetative tissues (Gong *et al.*, 2002;

Penterman *et al.*, 2007; Ortega-Galisteo *et al.*, 2008). The *ros1 dml2 dml3* triple mutant has no overt morphological phenotype and results in increased methylation at limited loci without affecting the global methylation status (Penterman *et al.*, 2007). The maternal *dme* mutation confers aborted seeds, i.e., the maternal *dme* allele fails to be transmitted to progeny (Choi *et al.*, 2002, 2004). In addition, in the endosperm from seeds with a maternal *dme* allele (*dme* endosperm) characteristic global non-CG hypomethylation and local CG hypermethylation occur as compared with methylation level in the wild-type endosperm (Gehring *et al.*, 2009a; Hsieh *et al.*, 2009). Moreover, *DME* was recently shown to be expressed in the vegetative cell of the male gametophyte, even though the level of *DME* mRNA in pollen is extremely lower than in ovule, and required for demethylation of some imprinted genes and transposons (Schoft *et al.*, 2011). They also showed that reduced transmission of *dme* alleles was observed in two Col ecotypes, Col-*gl and* Col-0, although *dme* alleles were efficiently transmitted in the *Ler* ecotype.

Phylogenetic analysis revealed that the rice genome has six putative bifunctional DNA glycosylases that mediate cytosine DNA demethylation, four *ROS1* orthologs and two *DML3* orthologs, but no *DME* orthologs (Zemach *et al.*, 2010). Rice endosperm DNA is hypomethylated in all sequence contexts, implying that hypomethylation in rice endosperm must rely on some of these DNA

glycosylases or alternate biochemical mechanisms (Zemach *et al.*, 2010). In this study, to characterize the function of one of the four rice *ROS1* orthologs, which is tentatively named *ROS1a* (LOC_Os01g11900.1) and which resides on chromosome 1, homologous recombination (HR)-promoted knock-in targeting with positive-negative selection was employed (Yamauchi *et al.*, 2009) to obtain a mutant that disrupts *ROS1a* by fusing its endogenous promoter with the *GUS* reporter gene for β-glucuronidase. T_0 plants with the null knock-in allele, *ros1a-GUS1*, in the heterozygous condition were reproducibly obtained and *GUS* expression was detected in the T_0 plants in the shoot apical, lateral, and inflorescence meristems as well as in both female and male gametophytes before fertilization. The *ros1a-GUS1* allele virtually failed to be transmitted to the next generation; neither the maternal nor the paternal *ros1a-GUS1* allele could be found in the progeny. The results indicate that *ROS1a* must be indispensable in both gametophytes, presumably via DNA demethylation, and that the null allele of *ROS1a* would be difficult to isolate by conventional mutagenesis techniques, in which mutants are usually obtained as segregants in the progeny population. The available data further indicate that *ROS1a* appears to play, to a certain extent, analogous roles to *DME* in both male and female gametophytes. Based on the results, I also discuss similarities and differences between *ROS1a* in rice and *DME* in Arabidopsis.

MATERIALS AND METHODS

Nucleic acid procedures

General nucleic acid procedures, including plasmid preparation, plant DNA and RNA preparation, PCR and RT-PCR amplification, and Southern blot and DNA sequencing analyses were performed as described (Terada *et al.*, 2007; Johzuka-Hisatomi *et al.*, 2008). qRT-PCR was performed as described by Yamauchi *et al.* (2009). The sequences of primers are listed (Table 5).

The cDNA ends were cloned by using 5'- and 3'-RACE performed with the Gene RACER kit (Invitrogen, Carlsbad, CA, USA, currently Lifetechnologies co.) using appropriate primers (Table 5). A full-length *ROS1a* cDNA was obtained by RT-PCR from total RNA isolated from Nipponbare by using appropriate primers (Table 5) and was verified by sequencing. Multiple amino-acid alignments were produced with a web-based version of ClustalW (http://clustalw.ddbj.nig.ac.jp/) using default settings.

Bacterial cell toxicity assay

The full-length *ROS1a* cDNA amplified with outside primers (Table 5) was inserted into the pMAL vector (New England BioLabs, Ipswich, MA, USA) to yield pMAL-ROS1a. To construct *ROS1a* cDNA containing the K1499A or D1515A point mutation, a simplified version of the assembly PCR method was used

(Stemmer *et al.*, 1995). Both 5' and 3' segments were amplified from the *ROS1a* cDNA with appropriate primers (Table 5), and full-length cDNA containing only the desired mutation was subsequently amplified with the outside primers and screened by DNA sequencing. The bacterial cell toxicity assay was performed as described by Gehring *et al.* (2006). A fresh colony was picked and resuspended in 5 ml of LB/Glu/Amp (LB supplemented with 0.2% [w/v] glucose and 100 μg/ml of ampicillin) liquid medium. After 12 h incubation at 37°C, the culture was diluted 200-, 1000-, 5000-, 25000-, and 125000-fold, and an aliquot (5 μl) was spotted on LB/Glu/Amp plates with or without IPTG (Sigma-Aldrich, St Louis, MO, USA). The plates were incubated at 30°C for 14 h.

Knock-in targeting

The vector used here for knock-in targeting, pJHY-Ki-ROS1a, is a derivative of the backbone of the knock-in targeting vector pJHY-Ki2 (Yamauchi *et al.*, 2009), which carries a 3.0-kb *Pac*I-*Srf*I fragment of the *ROS1a* promoter and a 3.0-kb *Pme*I-*Asc*I fragment of the 5' coding region of *ROS1a* (Figure 6b). The 3.0-kb fragments were prepared by PCR amplification of the Nipponbare genomic sequence with appropriate primers including the indicated restriction sites (Table 5) used for inserting into the corresponding cloning sites of pJHY-Ki2. Sequencing of the cloned fragments showed no base changes. The control plasmid

pJHY-Ki-ROS1aC, a pJHY-Ki-ROS1a derivative that includes a 3.5-kb *ROS1a* promoter fragment and a 3.5-kb fragment from the 5' *ROS1a* coding region to detect authentic 5'-JF and 3'-JF (Figure 6c), was constructed in the same way by cloning the 5'-flanking 0.5-kb I-*Ceu*I-*Pac*I and 3'-flanking 0.5-kb *Asc*I-I-*Sce*I fragments of the inserted 3.0-kb *Pac*I-*Srf*I and 3.0-kb *Pme*I-*Asc*I fragments, respectively. *Agrobacterium*-mediated rice transformation and screening for targeted calli were performed and characterized as described (Terada *et al.*, 2007; Yamauchi *et al.*, 2009). Histochemical GUS staining was performed as described (Yamauchi *et al.*, 2009). For differential interference contrast micrographs of ovary tissues, GUS-stained samples were passed through a series of methyl salicylate and ethanol mixture (ratios 1:2, 1:1, and 2:1), 1 h in each mixture, and finally 100 % methyl salicylate overnight. Immunostaining with anti-GUS antibody (Molecular Probe, Invitrogen, currently Lifetechnologies Co.) was performed according to the supplier's recommendation

Reciprocal crosses between Nipponbare and the knock-in T_0 plants

Cross-pollination in greenhouse was performed manually. One day before pollination, the top of each female-parent spikelet in a panicle was cut off using small scissors, and then all anthers were removed by grabbing their filaments and pulling gently with sharp forceps. The remaining florets, if present in the same

panicle, were also removed. The emasculated panicle was covered with a bag to avoid undesirable cross-pollination. Next day, when the male-parent flower was opened, its anthers were removed with forceps and put into the emasculated spikelet. The cross-pollinated spikelet was again covered with a bag to avoid cross-contamination.

Characterization of pollen

For staining with DAPI, pollen grains were fixed in an ethanol/acetic acid (3:1) solution for 1 h at room temperature, rehydrated through on ethanol series (75, 55, and 35%) for 5 min each, and stained with 1 μg/ml DAPI solution for 30 min at room temperature. Stained pollen was subsequently monitored with a fluorescence Biozero (BZ)-8000 microscope (KEYENCE, Osaka, Osaka, Japan) with BZ filter DAPI-BP. To evaluate pollen viability, anthers were removed before flowering and placed on a glass slide with 50 mM phosphate buffer (pH 7.0). The anthers were opened with a razor blade; approximately one-half of the pollens of each anther was used for iodine staining (Terada *et al.*, 2002) to examine for pollen viability (Figure 12), and the remaining half was stained for GUS expression to determine the segregation of the knock-in allele. To examine *in vitro* pollen germination, pollen grains from dehisced anthers were placed on a layer of cellophane on the pollen germination medium (13.5% [w/v] sucrose, 10 mM $CaCl_2$,

0.4 mM boric acid, 1 mM KCl, 0.001% [w/v] myo-inositol, and 0.6% [w/v] agarose). The released pollen was incubated at 25°C for 30 min in the semi-dark. Germinated pollen grains were then stained for GUS expression at 37°C for 6 h and observed with a microscope (KEYENCE) using bright-field illumination.

RESULTS

Characterization of *ROS1a*

Rice contains four *ROS1* orthologs and two *DML3* orthologs (Zemach *et al.*, 2010), which are tentatively named *ROS1a* to *ROS1d* and *DML3a* and *DML3b*, respectively, and contain the characteristic DNA glycosylase domains flanked by conserved domains of unknown functions (Figures 1a,b, 2, 3, and 4). Of these, *ROS1a* is the longest gene, comprising 17 exons that encode a protein with 1952 amino acids and 5'- and 3'-untranslated regions of 73 bp and 607 bp, respectively (Figure 1a). RT-PCR analysis revealed that *ROS1a* was expressed in all the vegetative and reproductive tissues tested (Figure 1c). Further quantitative RT-PCR (qRT-PCR) analysis revealed that *ROS1a* is the most extensively expressed gene among the four expressing genes, *ROS1a, ROS1c, ROS1d,* and *DML3a,* in the selected tissues examined including anthers and pistils, whereas *ROS1b* and *DML3b* are scarcely expressed in these tissues (Figure 4). Interestingly, considerable amounts of transcripts from *ROS1c, ROS1d,* and *DML3a* were detected in pistils and immature seeds two days after pollination. Public expression database (Rice Gene Expression [http://rice.plantbiology.msu.edu/expression.shtml]) conform to the results.

Arabidopsis *DME* encodes the bifunctional DNA glycosylase/lyase for 5-meC DNA demethylation, and the induced DME protein is toxic to *E. coli* strains

containing 5-meC in their genome, probably because of the formation of deleterious abasic sites following the excision of 5-meC and/or nicks in the *E. coli* genome (Gehring *et al.*, 2006). Moreover, either the substitution of the invariant lysine at position 1286 to glutamine (K1286Q) or the aspartic acid at position 1304 to asparagine (D1304N) within the DNA glycosylase domain of DME (Figure 1b) reduces DNA glycosylase activity, and inactive DME (D1304N) and active DME are nontoxic to *E. coli* strains with and without 5-meC, respectively. Whether induced rice ROS1a is toxic to *E. coli* containing 5-meC in their genome was examined. When *ROS1a* cDNA driven by an IPTG (isopropyl-β-D-thiogalactopyranoside)-inducible promoter was used, ROS1a was toxic to an *E. coli dcm*$^+$ strain with 5-meC in an IPTG-dependent manner and was less toxic to a *dcm*$^-$ mutant without 5-meC (Figure 5). The invariant lysine (K1499) and the aspartic acid (D1515) residues in rice ROS1a, which correspond to the invariant lysine (K1286) and aspartic acid (D1304) residues in Arabidopsis DME (Figure 1b), were modified to alanines (K1499A and D1515A, respectively) and examined. Expression of these K1499A and D1515A mutant cDNAs had little effect on either the *dcm*$^+$ or *dcm*$^-$ strains (Figure 5). Parallel experiments were also performed using the active DME and inactive DME (D1304N) cDNAs from Arabidopsis (Gehring *et al.*, 2006), and comparable results were obtained. These results are consistent with the notion that the rice ROS1a protein bears a

bifunctional DNA glycosylase/lyase for 5-meC DNA demethylation, although the biochemical characterization of the ROS1a enzyme remains to be achieved to confirm the notion.

Knock-in targeting of *ROS1a*

To generate a null mutation of *ROS1a* and to monitor the spatiotemporal expression of *ROS1a* in rice, knock-in targeting in which the endogenous *ROS1a* promoter was fused to the *GUS* reporter gene (Figure 6) was carried out, following the experimental design for knock-in targeting of *MET1a* (Yamauchi *et al.*, 2009). By identifying and subsequently sequencing the PCR-amplified junction fragments, which were generated by HR between the homologous segments in the introduced vector and the endogenous *ROS1a* gene, six independently targeted calli were obtained with a frequency of approximately 1.1% per surviving callus with positive-negative selection. Through multiple shoots, more than ten independent transgenic T_0 plants, comprising a plant line, were regenerated from each callus line obtained and subjected to Southern blot analysis (Figure 7). All of the T_0 plants tested were heterozygotes with only one copy of *GUS* integrated into the expected *ROS1a* site (*ROS1a/ros1a-GUS1*) and no additional ectopic events were detectable (Figure 7). Although these T_0 plants exhibited no overt morphological phenotypes during the vegetative phase, the *ROS1a* transcripts

that accumulated in three independently obtained lines were reproducibly about half as abundant as those in Nipponbare (Figure 8a), suggesting that *ROS1a* was expressed in a dosage-dependent manner.

Because the reproducible, dosage-dependent, and spatiotemporal expression of *GUS* was observed in transgenic plants after targeted *MET1a* knock-in (Yamauchi *et al.*, 2009), similar reproducible spatiotemporal expression of *GUS* was also anticipated in the heterozygous knock-in T_0 plants. Indeed, independently isolated *ROS1a/ros1a-GUS1* plants displayed very similar *GUS* expression patterns. GUS staining was detected in apical and lateral shoot meristems, inflorescence meristems, lodicules, and pollen (Figure 9a,b). Approximately one-half of both mature pollen grains and non-fertilized ovules of the *ROS1a/ros1a-GUS1* plants also displayed GUS-positive staining (Figure 9c,d), indicating that *GUS* of the *ros1a-GUS1* allele was expressed in these gametophytes. In the ovule, GUS-positive cells appeared to be antipodals, central cell, and egg apparatus comprising an egg cell and two synergids (Figure 9e-h), although it is hard to specify the synergids in the rice egg apparatus. The results are consistent with those from RT-PCR analysis, which showed endogenous *ROS1a* expression in male gametophyte-producing anthers and in female gametophyte-producing pistils (Figures 1c and 4). Although GUS staining was not detectable in elongated roots, root meristems, or adult leaves (Figure 8b),

RT-PCR analysis revealed that the expression of *ROS1a* and/or *GUS* in both tissues was detectable in Nipponbare (*ROS1a/ROS1a*) and in the *ROS1a/ros1a-GUS1* plants (Figures 1c, 4, and 8c). Presumably, the activity of the endogenous *ROS1a* promoter fused with *GUS* in the *ROS1a/ros1a-GUS1* plants was too weak to produce GUS-positive patterns in elongated roots and adult leaves but was strong enough to lead to positive GUS staining in meristematic cells as well as in both female and male gametophytes before fertilization.

Rare transmission of the *ros1a-GUS1* allele to progeny

To obtain homozygous knock-in plants, the isolated T_0 plants were self-pollinated. Of the selected 250 fully grown seeds, 232 germinated, and the resulting T_1 seedlings were genotyped by PCR analysis. Surprisingly, only two were *ROS1a/ros1a-GUS1* plants, and the remaining 230 seedlings bore the homozygous wild-type (*ROS1a/ROS1a*) allele (Table 1). Moreover, the *ros1a-GUS1* allele in the two heterozygous T_1 plants failed to be transmitted to progeny, confirming that the *ros1a-GUS1* allele was rarely transmittable to progeny. Close inspection revealed that the T_0 panicles were comprised of three different grains: empty grains with only infertile flower remnants (Figure 10b), grains with normal-shaped seeds (Figures 11a,b and 10c), and grains with deformed seeds containing severely underdeveloped and non–starch producing

endosperm (Figures 11c,d and 10d-n). Approximately equal numbers of the normal-shaped and deformed T_1 seeds were observed (54:56, 1:1, $\chi^2 = 0.036$, $P > 0.8$; Table 2). Embryos in the deformed seeds always displayed GUS-positive staining, whereas none of the normal-shaped seeds showed any GUS-positive patterns, indicating that the *ros1a-GUS1* allele co-segregated with the deformed seed phenotype and that the normal-shaped seeds were *ROS1a/ROS1a* (Table 2 and Figure 11b,d). In the analysis of the segregation ratio of *ros1a-GUS1* in seedlings (Table 1), the deformed seeds bearing the *ros1a-GUS1* allele were probably excluded from the analysis because only fully-grown seeds were used.

The normal-shaped and deformed seeds at 5 days after pollination (DAP) were also dissected. Like Nipponbare, the normal-shaped seeds had endosperm and a central cell cavity with discrete cells and a normal-shaped embryo composed of shoot and root meristems together with scutellum (Figure 11e,f). In contrast, the endosperm of the deformed seeds failed to develop at an early stage, and a large central cavity and a parent-derived thick nucellus that collapses during normal endosperm development were still present (Figure 11g,h). Those embryos developed more slowly and took on a slightly irregular shape as compared with wild-type embryos. Indeed, embryos in the deformed seeds displayed variable morphologies (Figures 11c,d and 10) ranging from no obvious distinctive organs in severely defective embryos (Figure 10e-h) to less defective embryos with

defective shoot or root meristems or multiple meristems (Figure 10i-o). The least defective embryos had apparently normally developed radicle, coleoptile, shoot apex, and thick scutellum (Figure 11d). The proportion of these embryo defects varied among the transgenic plants. Some of the least defective embryos could germinate on agar plates and grow normally once both the shoot and root were germinated, with the resulting plants having indistinguishable morphology and characteristic *ros1a-GUS1* segregation patterns as compared with their parent heterozygous plants (see Table 1).

Reciprocal crosses between Nipponbare and the knock-in T_0 plants

To characterize the rare transmission of the *ros1a-GUS1* allele further, reciprocal crosses between Nipponbare and the knock-in T_0 plants were carried out. Emasculated *ROS1a/ros1a-GUS1* flowers artificially pollinated with Nipponbare pollen produced both normal-shaped and deformed seeds (Table 3), all of which displayed GUS-negative and GUS-positive embryos coincident with normal and severely underdeveloped endosperms, respectively. These results are comparable to those of self-pollinated T_0 plants (Table 2) and indicate nonequivalent function of *ROS1a* with respect to maternal and paternal contributions to endosperm development, whereby only the maternal *ROS1a* allele was essential in endosperm development. Reciprocal crosses between

Nipponbare ovules and *ROS1a/ros1a-GUS1* pollen yielded only GUS-negative normal-shaped seeds (Table 3). Since most of the normal-shaped seeds, obtained from self-pollinated *ROS1a/ros1a-GUS1* plants, germinated and the resulting seedlings bore the *ROS1a/ROS1a* genotype (Table 1), the paternal *ros1a-GUS1* allele appears to fail to be transmitted to progeny. Because the percentage of empty grains containing infertile flower remnants in self-pollinated T_0 plants (Table 2) was significantly lower than 50% (20.1% ± 7.2 [SD]; n = 5), it is unlikely that they were aborted after fertilization because of the paternal *ros1a-GUS1* allele. Presumably, failure to transmit the paternal *ros1a-GUS1* allele was due to a male gametophytic defect(s) prior to fertilization. Although the ratio of normal-shaped seeds to deformed seeds in crosses between ovules from the *ROS1a/ros1a-GUS1* plants and Nipponbare pollen (Table 3) was considerably lower than that for self-pollination (Table 2), the cause of the bias remains unknown.

Germination of pollen from the knock-in T_0 plants

The vegetative phase of *ROS1a/ros1a-GUS1* plants was phenotypically indistinguishable from Nipponbare; they developed flowers with normal pistils and stamens, obvious anther dehiscence, and self-pollination (Figure 9). Compared with Nipponbare, the T_0 mutants had a comparable percentage of iodine-stained

pollen indicative of pollen viability, and the number of GUS-positive and GUS-negative pollen grains in the *ROS1a/ros1a-GUS1* anthers was approximately equal as expected (n = 1414, 680:737, 1:1, χ^2 = 2.29, P > 0.1; Table 4 and Figure 12a,b). DAPI (4', 6-diamidino-2-phenylindole)-stained mature pollen showed two sperm cells and one vegetative nucleus in pollen with both the *ros1a-GUS1* allele and the *ROS1a* allele (Figure 12c,d).

To elucidate the transmission failure of the paternal *ros1a-GUS1* allele further, germination ability was examined by scoring GUS-positive and GUS-negative pollen grains germinating on pollen germination medium. Considerably fewer GUS-positive than GUS-negative pollen grains germinated in the T_0 anthers (Figure 13), indicating that the *ros1a-GUS1* pollen germinated less efficiently than did the *ROS1a* pollen. In contrast, both GUS-positive and GUS-negative pollen grains germinated equally well in another heterozygous knock-in T_0 plant, in which the endogenous promoter of *DDM1a* (LOC_Os09g127060.1) was fused to *GUS* , one of two rice *DDM1* genes (Figure 13; generation of this rice line to be published elsewhere). Clearly the germination capability of the *ros1a-GUS1* pollen was reduced but not completely abolished, suggesting that other impaired processes must be involved in the failure of the paternal *ros1a-GUS1* transmission.

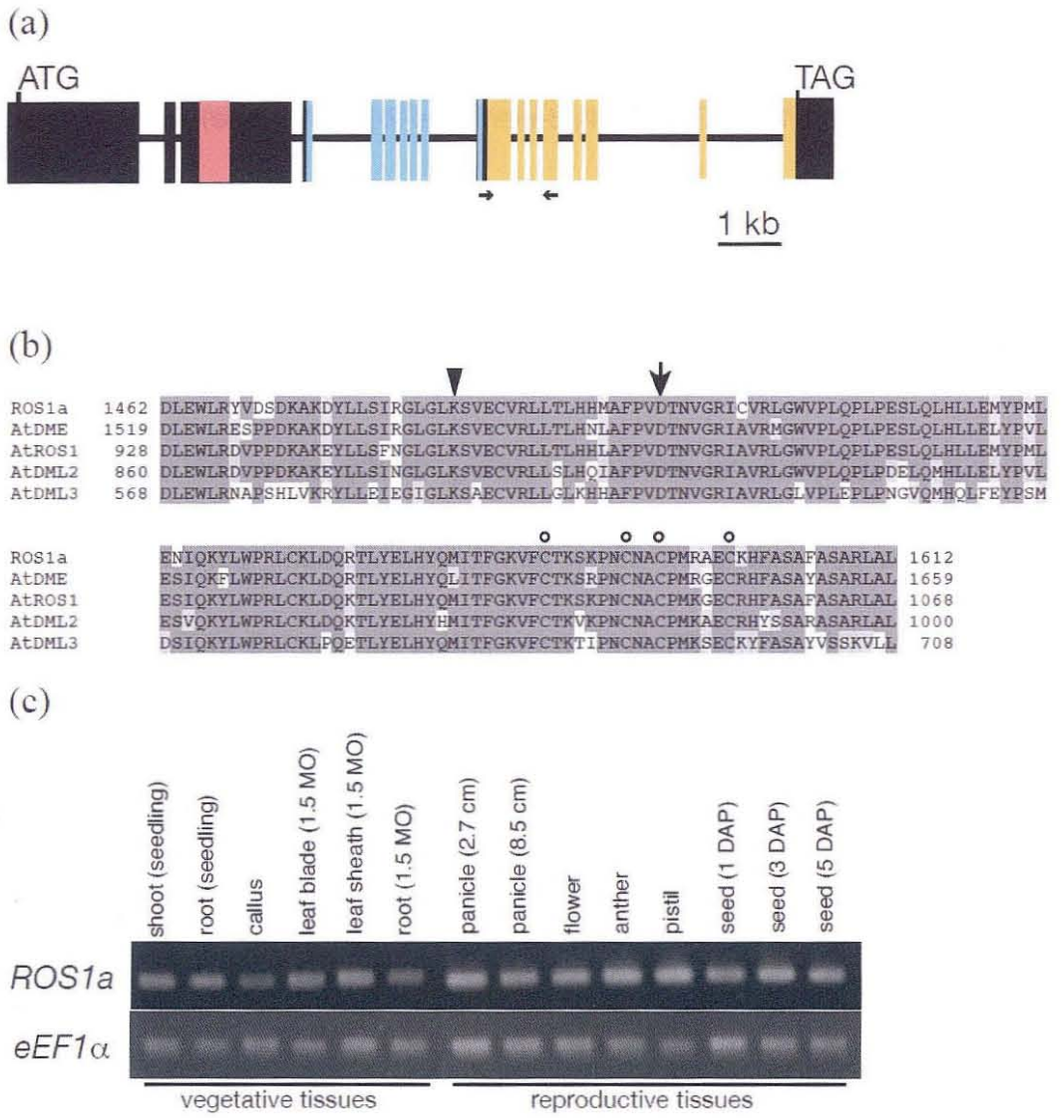

(a)
ATG
TAG
1 kb
(b)
ROS1a 1462 DLEWLRYVDSDKAKDYLLSIRGLGLKSVECVRLLTLHHMAFPVDTNVGRICVRLGWVPLQPLPESLQLHLLEMYPML
AtDME 1519 DLEWLRESPPDKAKDYLLSIRGLGLKSVECVRLLTLHNLAFPVDTNVGRIAVRMGWVPLQPLPESLQLHLLELYPVL
AtROS1 928 DLEWLRDVPPDKAKEYLLSFNGLGLKSVECVRLLTLHHLAFPVDTNVGRIAVRLGWVPLQPLPESLQLHLLEMYPML
AtDML2 860 DLEWLRDVPPDKAKEYLLSINGLGLKSVECVRLLSLHQIAFPVDTNVGRIAVRLGWVPLQPLPDELQMHLLELYPVL
AtDML3 568 DLEWLRNAPSHLVKRYLLEIEGIGLKSAECVRLLGLKHHAFPVDTNVGRIAVRLGLVPLEPLPNGVQMHQLFEYPSM
ROS1a ENIQKYLWPRLCKLDQRTLYELHYQMITFGKVFCTKSKPNCNACPMRAECKHFASAFASARLAL 1612
AtDME ESIQKFLWPRLCKLDQRTLYELHYQLITFGKVFCTKSRPNCNACPMRGECRHFASAYASARLAL 1659
AtROS1 ESIQKYLWPRLCKLDQKTLYELHYQMITFGKVFCTKSKPNCNACPMKGECRHFASAFASARLAL 1068
AtDML2 ESVQKYLWPRLCKLDQKTLYELHYHMITFGKVFCTKVKPNCNACPMKAECRHYSSARASARLAL 1000
AtDML3 DSIQKYLWPRLCKLPQETLYELHYQMITFGKVFCTKTIPNCNACPMKSECKYFASAYVSSKVLL 708
(c)
shoot (seedling)
root (seedling)
callus
leaf blade (1.5 MO)
leaf sheath (1.5 MO)
root (1.5 MO)
panicle (2.7 cm)
panicle (8.5 cm)
flower
anther
pistil
seed (1 DAP)
seed (3 DAP)
seed (5 DAP)
ROS1a
eEF1α
vegetative tissues
reproductive tissues

Figure 1

Figure 1. Organization and expression of *ROS1a*.

(a) Schematic representation of *ROS1a*. Boxes and bars indicate exons and introns, respectively. Blue exons encode the conserved DNA glycosylase domain, whereas pink and orange exons encode conserved domains of unknown function (Morales-Ruiz *et al.*, 2006). Arrows indicate the position of primers used for RT-PCR in (c).

(b) Amino acid sequence alignment of the conserved DNA glycosylase domain from ROS1a with the corresponding regions from AtDME, AtROS1, AtDML2, and AtDML3 of Arabidopsis showing amino acid positions (numerals), identical (dark gray) or similar (light gray) amino acids, the conserved lysine (arrowhead) and cysteine (open circles) residues, and the conserved aspartic acid residue at the active site in glycosylases/lyases (arrow).

(c) The expression of *ROS1a* detected by RT-PCR analysis in various tissues. MO, month-old; DAP, days after pollination.

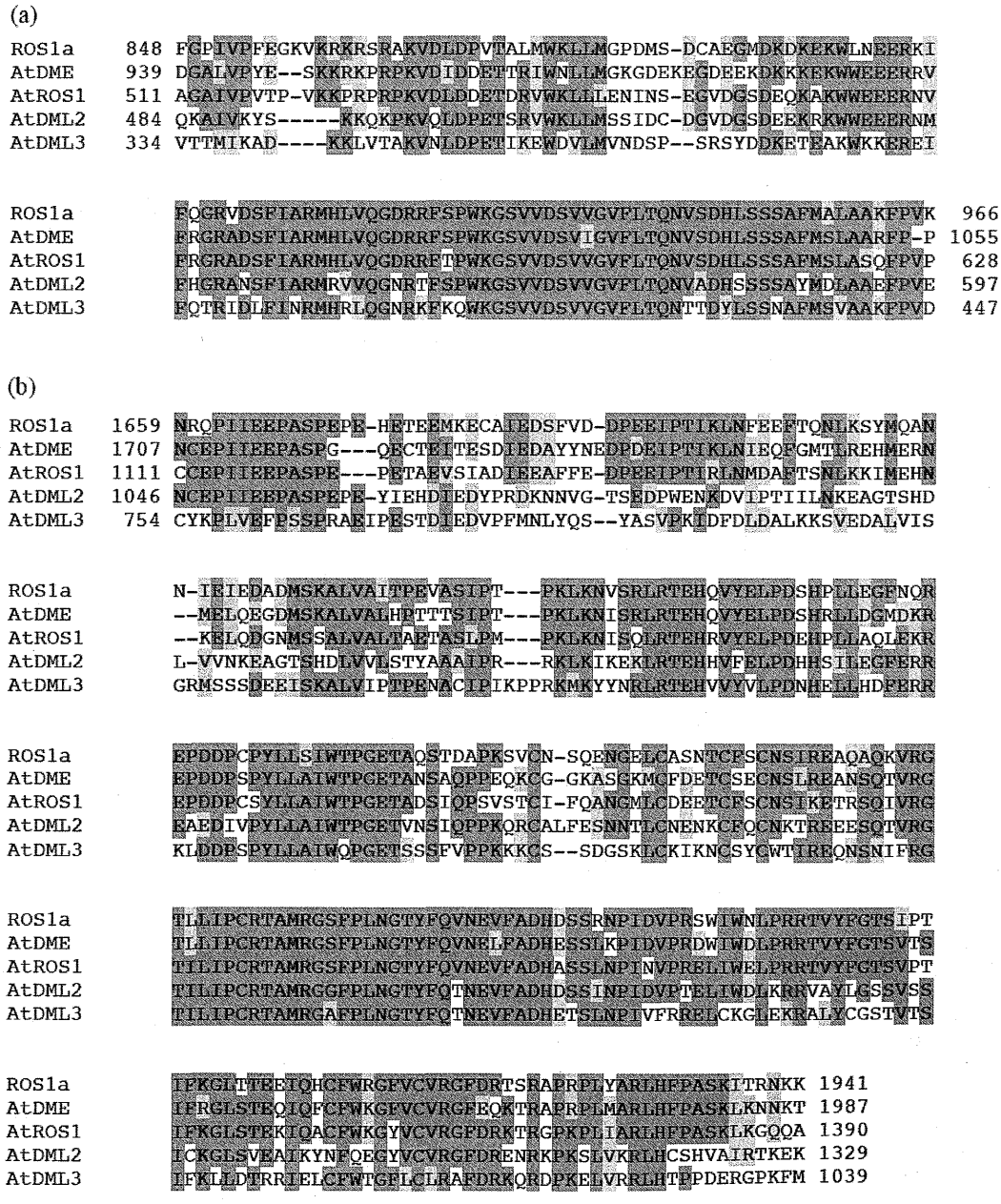

(a)
ROS1a 848 FGPIVPFEGKVKRKRSRAKVDLDPVTALMWKLLMGPDMS-DCAEGMDKDKEKWLNEERKI
AtDME 939 DGALVPYE--SKKRKPRPKVDIDDETTRIWNLLMGKGDEKEGDEEKDKKKEKWWEEERRV
AtROS1 511 AGAIVPVTP-VKKPRPRPKVDLDDETDRVWKLLLENINS-EGVDGSDEQKAKWWEEERNV
AtDML2 484 QKAIVKYS-----KKQKPKVQLDPETSRVWKLLMSSIDC-DGVDGSDEEKRKWWEEERNM
AtDML3 334 VTTMIKAD----KKLVTAKVNLDPETIKEWDVLMVNDSP--SRSYDDKETEAKWKKERET
ROS1a FQGRVDSFIARMHLVQGDRRFSPWKGSVVDSVVGVFLTQNVSDHLSSSAFMALAAKFPVK 966
AtDME FRGRADSFIARMHLVQGDRRFSPWKGSVVDSVIGVFLTQNVSDHLSSSAFMSLAARFP-P 1055
AtROS1 FRGRADSFIARMHLVQGDRRFTPWKGSVVDSVVGVFLTQNVSDHLSSSAFMSLASQFPVP 628
AtDML2 FHGRANSFIARMRVVQGNRTFSPWKGSVVDSVVGVFLTQNVADHSSSSAYMDLAAEFPVE 597
AtDML3 FQTRIDLFINRMHRLQGNRKFKQWKGSVVDSVVGVFLTQNTTDYLSSNAFMSVAAKFPVD 447
(b)
ROS1a 1659 NRQPIIEEPASPEPE-HETEEMKECAIEDSFVD-DPEEIPTIKLNFEEFTQNLKSYMQAN
AtDME 1707 NCEPIIEEPASPG---QECTEITESDIEDAYYNEDPDEIPTIKLNIEQFGMTTREHMERN
AtROS1 1111 CCEPIIEEPASPE---PETAEVSIADIEEAFFE-DPEEIPTIRLNMDAFTSNLKKIMEHN
AtDML2 1046 NCEPIIEEPASPEPE-YIEHDIEDYPRDKNNVG-TSEDPWENKDVIPTIILNKEAGTSHD
AtDML3 754 CYKPLVEFPSSPRAEIPESTDIEDVPFMNLYQS--YASVPKIDFDLDALKKSVEDALVIS
ROS1a N-IEIEDADMSKALVAITPEVASIPT---PKLKNVSRLRTEHQVYELPDSHPLLEGFNQR
AtDME --MELQEGDMSKALVALHPTTTSIPT---PKLKNISRLRTEHQVYELPDSHRLLDGMDKR
AtROS1 --KELQDGNMSSALVALTAPTASLPM---PKLKNISQLRTEHRVYELPDEHPLLAQLEKR
AtDML2 L-VVNKEAGTSHDLVVLSTYAAAIPR---RKLKIKEKLRTEHHVFELPDHHSILEGFERR
AtDML3 GRMSSSDEEISKALVIPTPENACIPIKPPRKMKYYNRLRTEHVVYVLPDNHELLHDFERR
ROS1a EPDDPCPYLLSIWTPGETAQSTDAPKSVCN-SQENGELCASNTCFSCNSIREAQAQKVRG
AtDME EPDDPSPYLLAIWTPGETANSAQPPEQKCG-GKASGKMCFDETCSECNSLREANSQTVRG
AtROS1 EPDDPCSYLLAIWTPGETADSIQPSVSTCI-FQANGMLCDEETCFSCNSIKETRSQIVRG
AtDML2 EAEDIVPYLLAIWTPGETVNSIQPPKQRCALFESNNTLCNENKCFQCNKTREEESQTVRG
AtDML3 KLDDPSPYLLAIWQPGETSSSFVPPKKKCS--SDGSKLCKIKNCSYGWTTREQNSNIFRG
ROS1a TLLIPCRTAMRGSFPLNGTYFQVNEVFADHDSSRNPIDVPRSWIWNLPRRTVYFGTSIPT
AtDME TLLIPCRTAMRGSFPLNGTYFQVNELFADHESSLKPIDVPRDWIWDLPRRTVYFGTSVTS
AtROS1 TILIPCRTAMRGSFPLNGTYFQVNEVFADHASSLNPINVPRELIWELPRRTVYFGTSVPT
AtDML2 TILIPCRTAMRGGFPLNGTYFQTNEVFADHDSSINPIDVPTELIWDLKRRVAYLGSSVSS
AtDML3 TILIPCRTAMRGAFPLNGTYFQTNEVFADHETSLNPIVFRRELCKGLEKRALYCGSTVTS
ROS1a IPKGLTTRETQHCFWRGFVCVRGFDRTSRAPRPLYARLHFPASKITRNKK 1941
AtDME IFRGLSTEQIQFCFWKGFVCVRGFEQKTRAPRPLMARLHFPASKLKNNKT 1987
AtROS1 IFKGLSTEKIQACFWKGYVCVRGFDRKTRGPKPLIARLHFPASKLKGQQA 1390
AtDML2 ICKGLSVEAIKYNFQEGYVCVRGFDRENRKPKSLVKRLHCSHVAIRTKEK 1329
AtDML3 IFKLLDTRRIELCFWTGFLCLRAFDRKQRDPKELVRRLHTPPDERGPKFM 1039

Figure 2

Figure 2. Amino acid sequence alignment of ROS1a with Arabidopsis AtDME, AtROS1, AtDML2, and AtDML3.

(a) Comparison of the 5' conserved domains (amino acid positions 848–966 of ROS1a).

(b) Comparison of the 3' conserved domains (amino acid positions 1659–1941 of ROS1a).

Numerals indicate amino acid positions. Dark and light grey backgrounds indicate identical and similar amino acids, respectively.

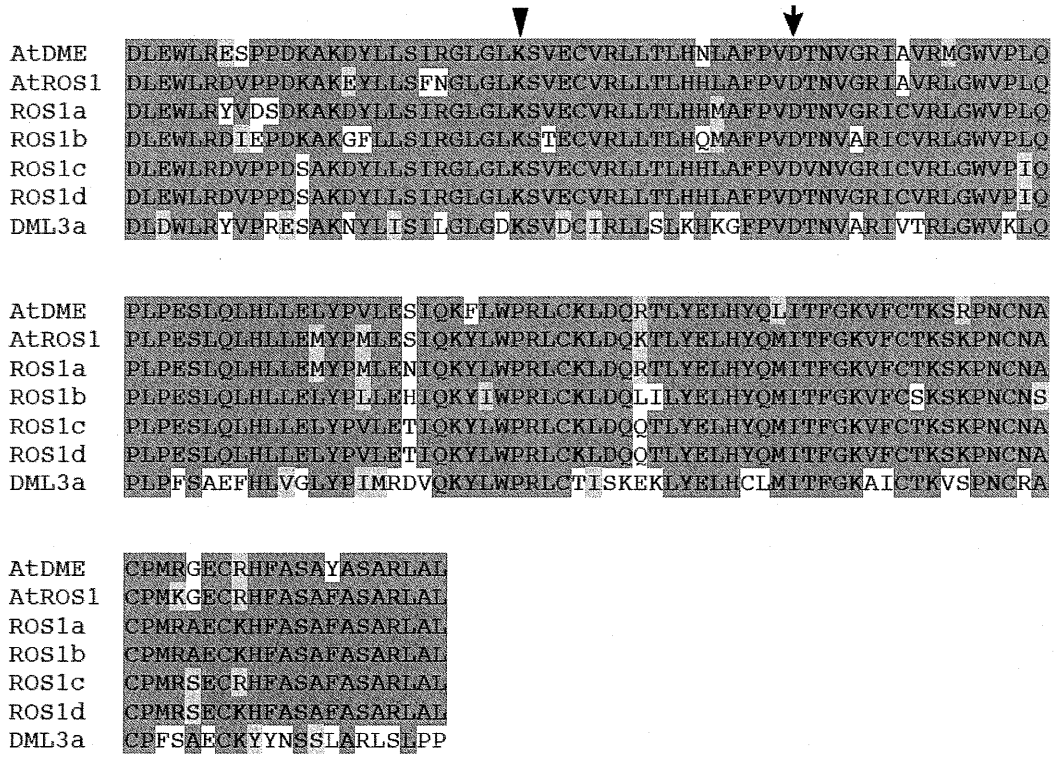

AtDME DLEWLRESPPDKAKDYLLSIRGLGLKSVECVRLLTLHNLAFPVDTNVGRIAVRMGWVPLQ
AtROS1 DLEWLRDVPPDKAKEYLLSFNGLGLKSVECVRLLTLHHLAFPVDTNVGRIAVRLGWVPLQ
ROS1a DLEWLRYVDSDKAKDYLLSIRGLGLKSVECVRLLTLHHMAFPVDTNVGRICVRLGWVPLQ
ROS1b DLEWLRDIEPDKAKGFLLSIRGLGLKSTECVRLLTLHQMAFPVDTNVARICVRLGWVPLQ
ROS1c DLEWLRDVPPDSAKDYLLSIRGLGLKSVECVRLLTLHHLAFPVDVNVGRICVRLGWVPIQ
ROS1d DLEWLRDVPPDSAKDYLLSIRGLGLKSVECVRLLTLHHLAFPVDTNVGRICVRLGWVPIQ
DML3a DLDWLRYVPRESAKNYLISILGLGDKSVDCIRLLSLKHKGFPVDTNVARIVTRLGWVKLQ
AtDME PLPESLQLHLLELYPVLESIQKFLWPRLCKLDQRTLYELHYQLITFGKVFCTKSRPNCNA
AtROS1 PLPESLQLHLLEMYPMLESIQKYLWPRLCKLDQKTLYELHYQMITFGKVFCTKSKPNCNA
ROS1a PLPESLQLHLLEMYPMLENIQKYLWPRLCKLDQRTLYELHYQMITFGKVFCTKSKPNCNA
ROS1b PLPESLQLHLLELYPLLEHIQKYIWPRLCKLDQLILYELHYQMITFGKVFCSKSKPNCNS
ROS1c PLPESLQLHLLELYPVLETIQKYLWPRLCKLDQQTLYELHYQMITFGKVFCTKSKPNCNA
ROS1d PLPESLQLHLLELYPVLETIQKYLWPRLCKLDQQTLYELHYQMITFGKVFCTKSKPNCNA
DML3a PLPFSAEFHLVGLYPIMRDVQKYLWPRLCTISKEKLYELHCLMITFGKAICTKVSPNCRA
AtDME CPMRGECRHFASAYASARLAL
AtROS1 CPMKGECRHFASAFASARLAL
ROS1a CPMRAECKHFASAFASARLAL
ROS1b CPMRAECKHFASAFASARLAL
ROS1c CPMRSECRHFASAFASARLAL
ROS1d CPMRSECKHFASAFASARLAL
DML3a CPFSAECKYYNSSLARLSLPP

Figure 3

Figure 3. Amino acid sequence alignment of five putative rice ROS1/DME3 glycosylase domains with the corresponding domains of Arabidopsis DME and ROS1.

The amino acid sequences of putative rice proteins were based on their transcripts available in public databases (MSU Rice Genome Annotation Release 7 [http://rice.plantbiology.msu.edu] or Rice Expression Profile Database Release 1.5 [http://ricexpro.dna.affrc.go.jp]): *ROS1b*, LOC_Os02g29230.1, seq-ID S-14907; *ROS1c*, LOC_Os05g37350.1, AK103871; *ROS1d*, LOC_Os05g37410.1, seq-ID S-14873; *DML3a*, LOC_Os02g29380.1, AK071406. The structure of the *ROS1a* transcript was also analyzed in this study (Figure 1a,b). Note that the public databases indicated that the remaining *DML3* gene, *DML3b* (LOC_Os04g28860.1), is unlikely to be expressed. Arrowhead and arrow indicate the conserved lysine and aspartic acid residues, respectively, and other symbols are as in Figure 2.

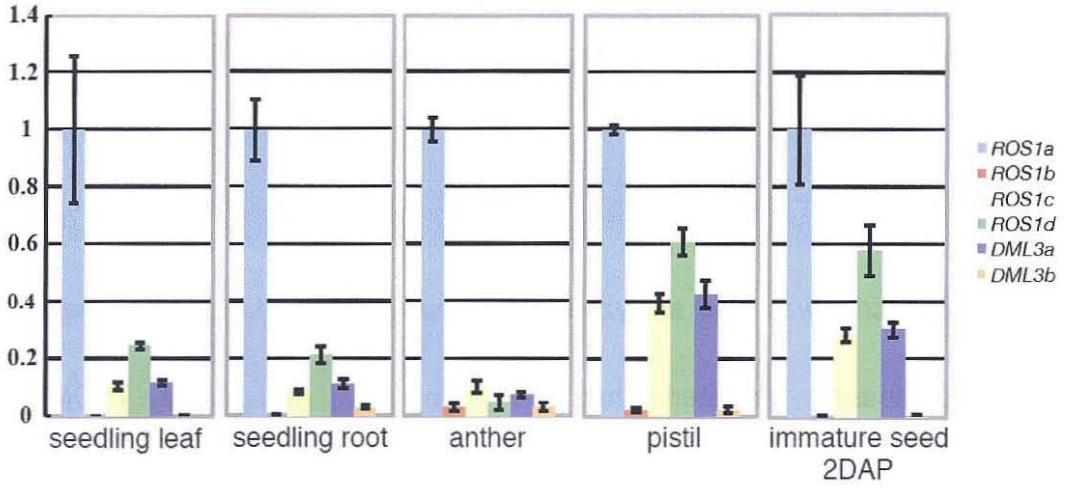

1.4
1.2
1
0.8
0.6
0.4
0.2
0
seedling leaf
seedling root
anther
pistil
immature seed 2DAP
ROS1a
ROS1b
ROS1c
ROS1d
DML3a
DML3b

Figure 4

Figure 4. Relative amounts of accumulated transcripts from *ROS1a, ROS1b, ROS1c, ROS1d, DML3a,* and *DML3b* in selected tissues by qRT-PCR analysis. The expression of the indicated genes relative to *Actin* was measured in the tissues indicated. The average of three observed values from *ROS1a* in each tissue was set to 1.0. Error bars represent the SD (n=3). Note that the efficiencies of the different primer sets used for qRT-PCR analysis (Table 5) were found to be similar. The primes for *DML3b* were designed based on MSU Rice Genome Annotation Release 7 (see Figure 3).

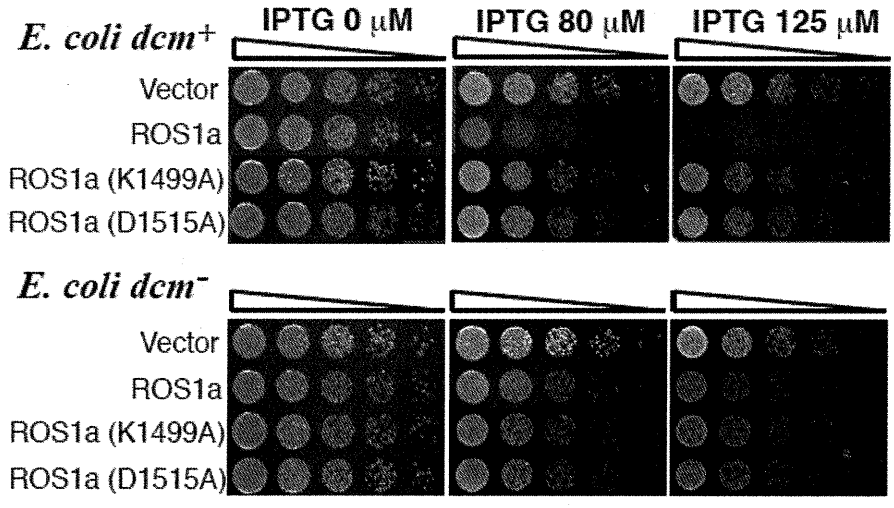

E. coli dcm+
IPTG 0 μM
IPTG 80 μM
IPTG 125 μM
Vector
ROS1a
ROS1a (K1499A)
ROS1a (D1515A)
E. coli dcm-
Vector
ROS1a
ROS1a (K1499A)
ROS1a (D1515A)

Figure 5

Figure 5. The toxic effects of ROS1a in *E. coli* containing 5-meC in the genome. *E. coli* K12 strains GM30 *dcm*$^+$ and GM31 *dcm*$^-$ (Palmer and Marinus, 1994) were used as host bacteria. Vector, ROS1a, ROS1a (K1499A), and ROS1a (D1515A) indicate the introduced vectors pMAL, pMAL-ROS1a, and pMAL-ROS1a derivatives containing the K1499A and D1515A point mutations, respectively. The bacterial culture was diluted and spotted from left to right.

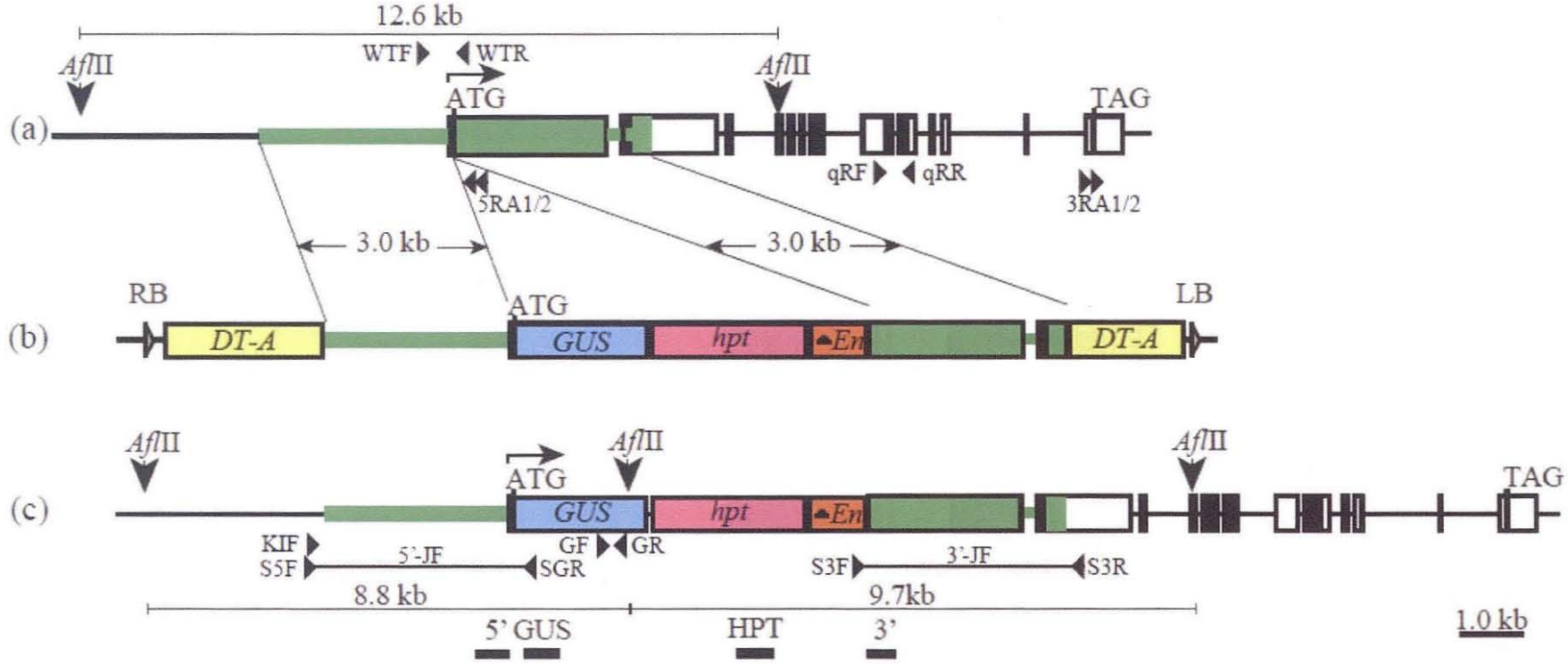

12.6 kb
WTF
WTR
AflII
ATG
AflII
TAG
(a)
qRF
qRR
5RA1/2
3RA1/2
3.0 kb
3.0 kb
RB
ATG
LB
(b)
DT-A
GUS
hpt
En
DT-A
AflII
ATG
AflII
AflII
TAG
(c)
GUS
hpt
En
KIF
S5F
5'-JF
SGR
GF
GR
S3F
3'-JF
S3R
8.8 kb
9.7kb
1.0 kb
5' GUS
HPT
3'

Figure 6

Figure 6. Strategy for the knock-in modification of *ROS1a* in rice.

(a) Genomic structure of *ROS1a*. Boxes and horizontal arrow indicate *ROS1a* exons and transcriptional initiation, respectively. Green segments are also carried by the vector pJHY-Ki-ROS1a. Arrowheads with 5RA1/2 and 3RA1/2 indicate primers for 5'- and 3'-RACE analysis, respectively, and those with qRF/qRR indicate primers for RT-qPCR analysis of *ROS1a*.

(b) T-DNA region of pJHY-Ki-ROS1a. RB, right border; *DT-A*, *diphtheria toxin A fragment* gene for negative selection; *GUS*, *β-glucuronidase* gene; *hpt, hygromycin phosphotransferase* gene for hygromycin B resistance; *ΔEn,* the 3' end of the maize *En* element; LB, left border.

(c) Genomic structure of the *ROS1a* locus targeted for knock-in. The 5'- and 3'-JFs indicate 5'- and 3'-junction fragments generated by HR between homologous segments in *ROS1a* and pJHY-Ki-ROS1a, respectively. Flanking arrowheads S5F/SGR and S3F/S3R represent primers for PCR analysis. Arrowheads with GF/GR indicate primers for RT-PCR analysis of *GUS*. Primers WTF/WTR and KIF/SGR were used for PCR genotyping of the *ROS1a* or *ros1a-GUS1* alleles to detect the 0.9- or 3.5-kb fragment, respectively. Horizontal thick-black bars with 5', GUS, HPT, and 3' indicate DNA probes for Southern blot hybridization. The sizes of *Afl*II restriction fragments are also indicated.

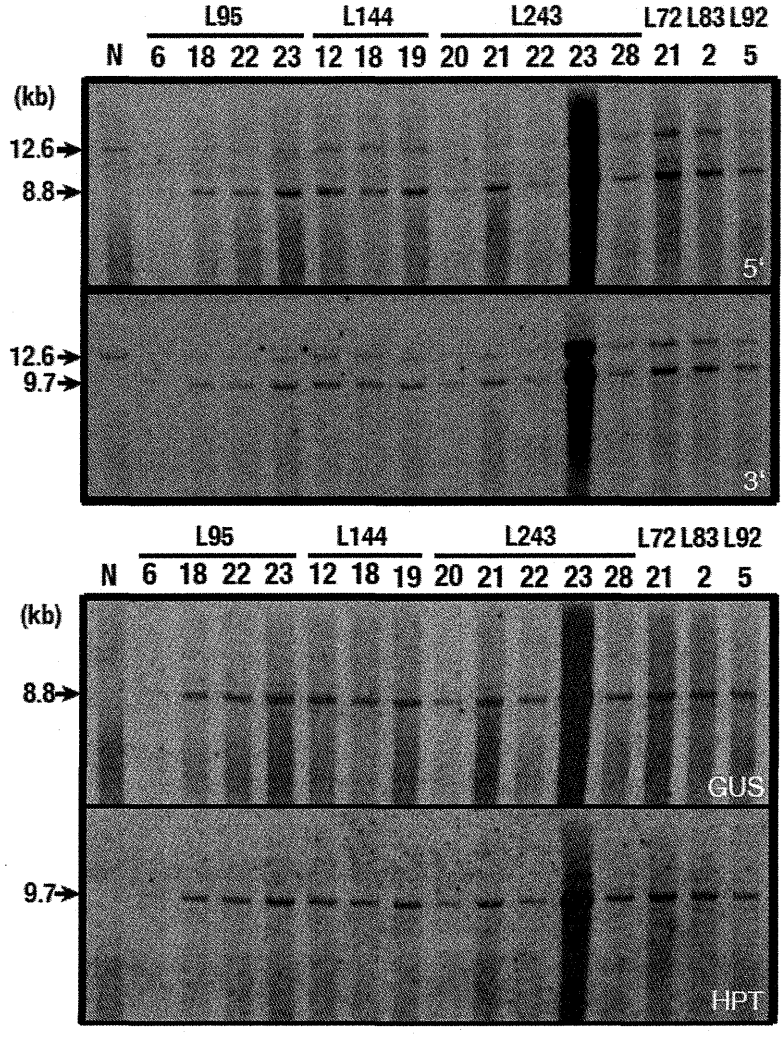

L95
L144
L243
L72 L83 L92
N 6 18 22 23 12 18 19 20 21 22 23 28 21 2 5
(kb)
12.6
8.8
5'
12.6
9.7
3'
L95
L144
L243
L72 L83 L92
N 6 18 22 23 12 18 19 20 21 22 23 28 21 2 5
(kb)
8.8
GUS
9.7
HPT

Figure 7

Figure 7. Southern blot analysis of the locus targeted for knock-in.

The *Afl*II fragments of genomic DNA samples prepared from leaves of T_0 plants from six plant lines (designated as L95, L144, L243, L72, L83, and L92) and those from the control Nipponbare plant (N) were hybridized to the probes shown in Figure 6c. Note that numerals represent different plants from each indicated line.

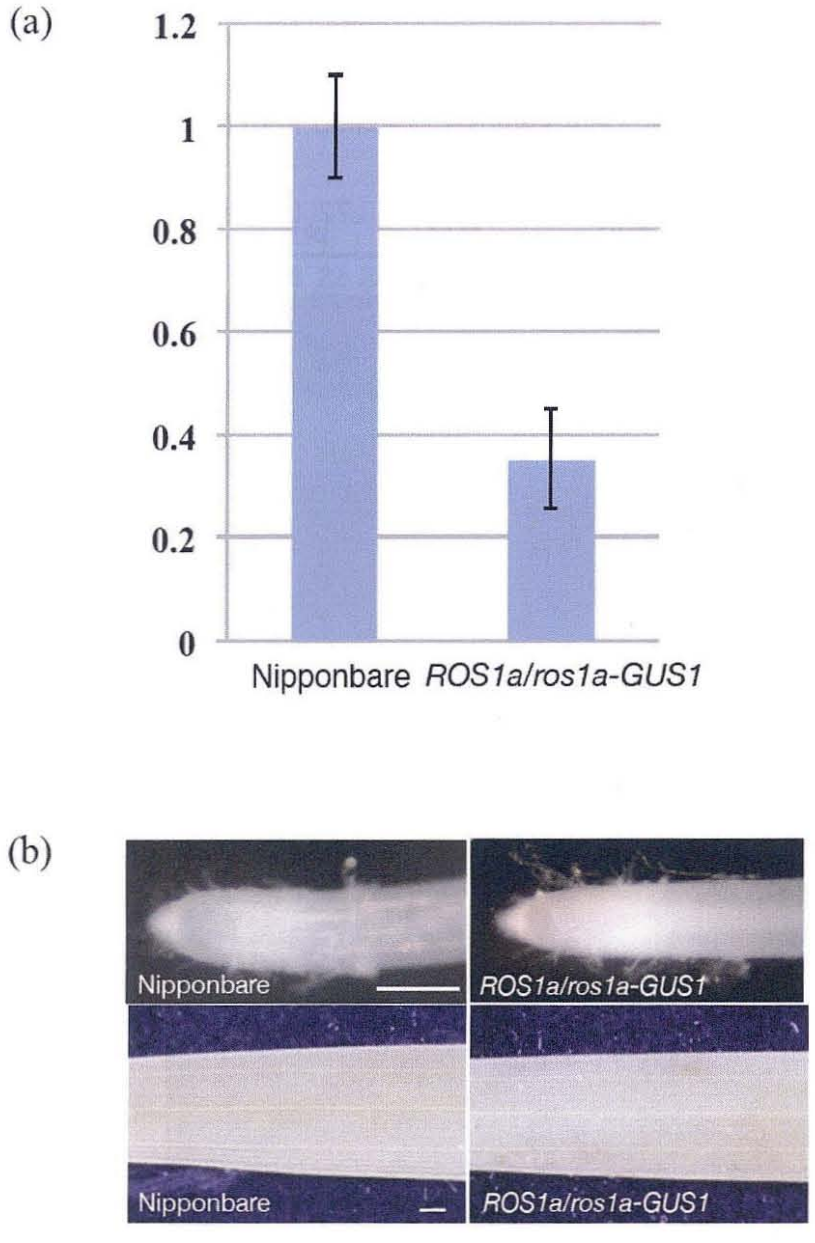

(a)
1.2
1
0.8
0.6
0.4
0.2
0
Nipponbare
ROS1a/ros1a-GUS1
(b)
Nipponbare
ROS1a/ros1a-GUS1
Nipponbare
ROS1a/ros1a-GUS1
(c)
root
adult leaf
L95-23
L144-12
L243-21
L95-23
L144-12
L243-21
Nipponbare
GUS +RT
-RT
eEF1α +RT
-RT

Figure 8

Figure 8. Expression of *ROS1a* or *GUS* in *ROS1a/ros1a-GUS1* plants.

(a) Expression of *ROS1a* in *ROS1a/ros1a-GUS1* plants. qRT-PCR analysis of the *ROS1a* mRNA in mature leaves was used to measure the expression of *ROS1a* relative to *Actin*. The average of three observed values from Nipponbare plants was set to 1.0. Error bars represent the SD (n = 3).

(b) Histochemical GUS staining in elongated roots including root meristems and adult leaves of *ROS1a/ros1a-GUS1* plants. Bars = 1 mm.

(c) The expression of *GUS* detected by RT-PCR analysis in elongated roots and adult leaves of *ROS1a/ros1a-GUS1* plants. Samples were prepared from three *ROS1a/ros1a-GUS1* plants (L95-23, L144-12, and L243-21) and Nipponbare. RT-PCR analysis was performed as in Figure 1c.

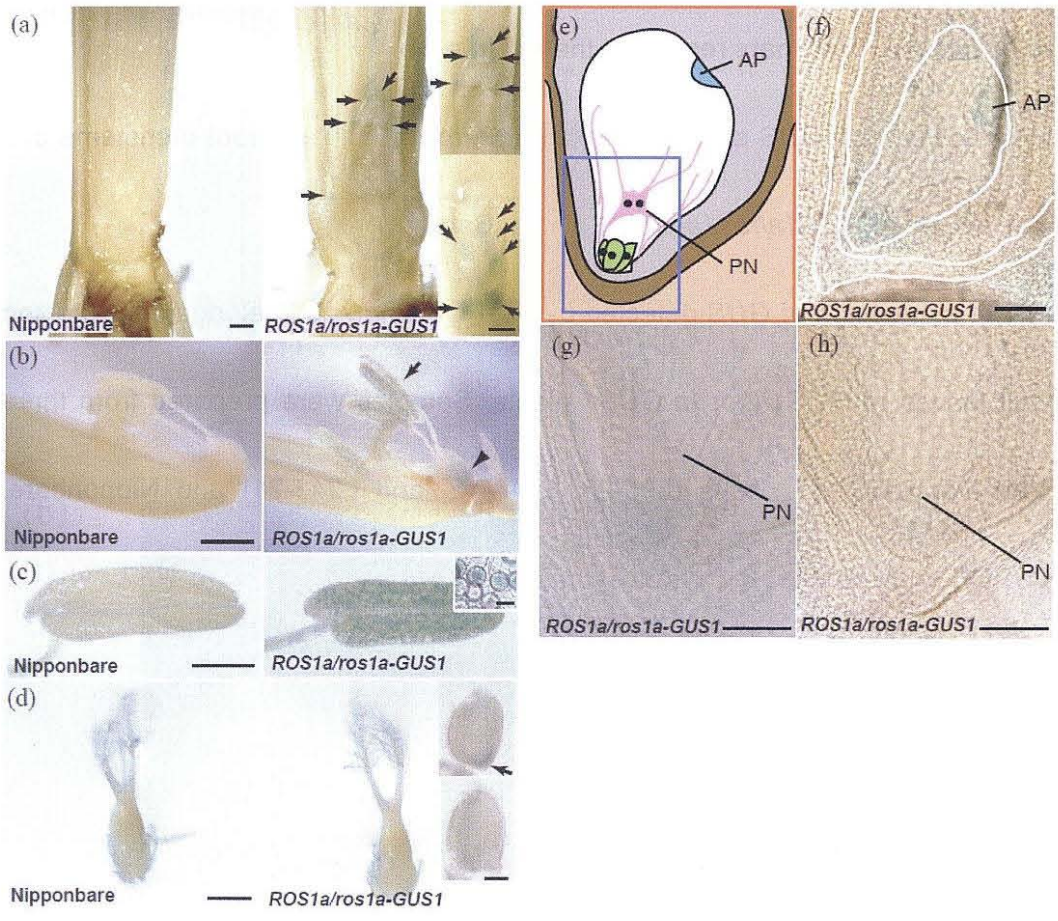

(a)
Nipponbare
ROS1a/ros1a-GUS1
(b)
Nipponbare
ROS1a/ros1a-GUS1
(c)
Nipponbare
ROS1a/ros1a-GUS1
(d)
Nipponbare
ROS1a/ros1a-GUS1
(e)
AP
PN
(f)
AP
ROS1a/ros1a-GUS1
(g)
PN
ROS1a/ros1a-GUS1
(h)
PN
ROS1a/ros1a-GUS1

Figure 9

Figure 9. Histochemical GUS staining patterns in *ROS1a/ros1a-GUS1* plants.

(a) Longitudinal cut through basal shoot meristem region of mature plants. Insets are magnified views of shoot apical and lateral (upper) and inflorescence (lower) meristems.

(b) Flowers.

(c) Anthers. Approximately half of the pollen present showed GUS staining (see also Figure 12b). Inset is a magnified view.

(d) Pistils. Insets are ovules before pollination that were exposed after removing the carpel. Upper and lower panels show a GUS-stained ovule and its sibling, which displayed no GUS staining, respectively. Before pollination, ovules containing the gametes with *ROS1a* were morphologically indistinguishable from those with *ros1a-GUS1*.

(e) Schematic representation of a rice female gametophyte enclosed by the maternal tissues of ovule; ovary tissue (light brown), integument (dark brown), and nucellus (gray). The haploid female gametophyte consists of egg apparatus (green), including egg cell and two synergids, central cell (white), comprising large vacuoles and thin lines of cytoplasm (pink), and antipodals (blue).

(f,g) Differential interference contrast micrographs of GUS-stained female gametophytes of *ROS1a/ros1a-GUS1* plants. To make the border clear, white lines are drawn into the image (f), which corresponds to the area enclosed by the

red line in (e).

(h) A differential interference contrast micrograph of its sibling female gametophyte displaying no GUS staining.

Arrows indicate GUS-stained meristems (a), pollen (b), and the likely egg apparatus (d). Arrowhead indicates the GUS-stained lodicule (b). AP, antipodal; PN, polar nuclei. The images in (g,h) correspond to the area enclosed by the blue line in (e). Bars = 1 mm in (a) and (b) and inset in (a), 500 μm in (c) and (d), 100 μm in inset in (d), and 50 μm in inset in (c) and in (f,g).

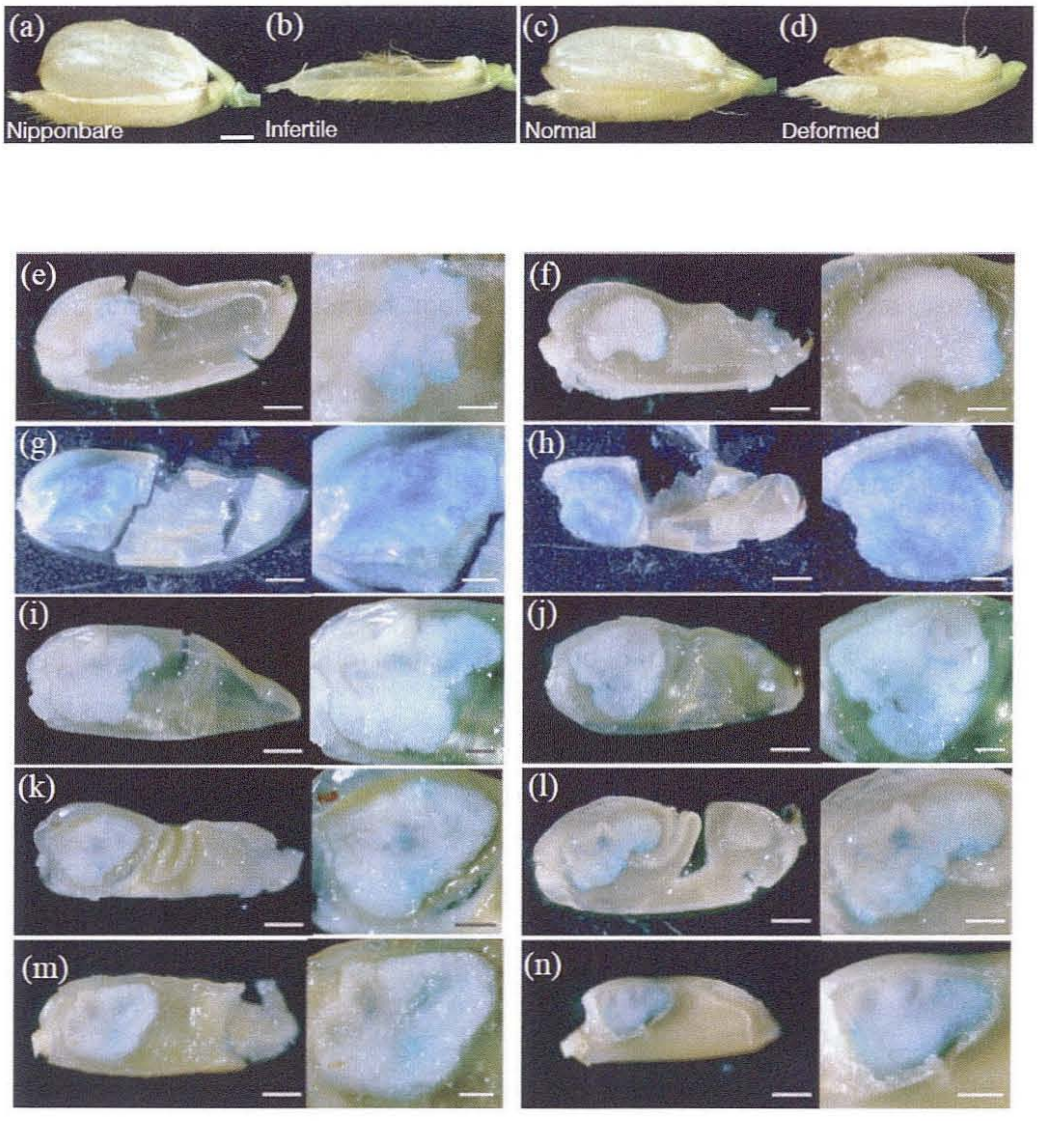


Figure 10

Figure 10. Various aberrant embryo phenotypes in deformed seeds.

Top panels (a-d) indicate a normal grain of the control plant Nipponbare (a), a typical empty grain with only an infertile flower remnant (b, Infertile), a grain with normal-shaped seed (c, Normal), and a grain with deformed seeds containing severely underdeveloped endosperm (d, Deformed) of self-pollinated *ROS1a/ros1a-GUS1* plants. Lemmas were removed to clearly show the phenotypes. The middle panels (e-n) display deformed seeds that were cut into halves and stained for GUS expression, with magnified views of embryos on the right side. Bars = 1 mm (left) and 500 μm (right). The bottom panel (o) shows a longitudinal section of a deformed seed. Arrowheads indicate meristematic structures. Bar = 100 μm.

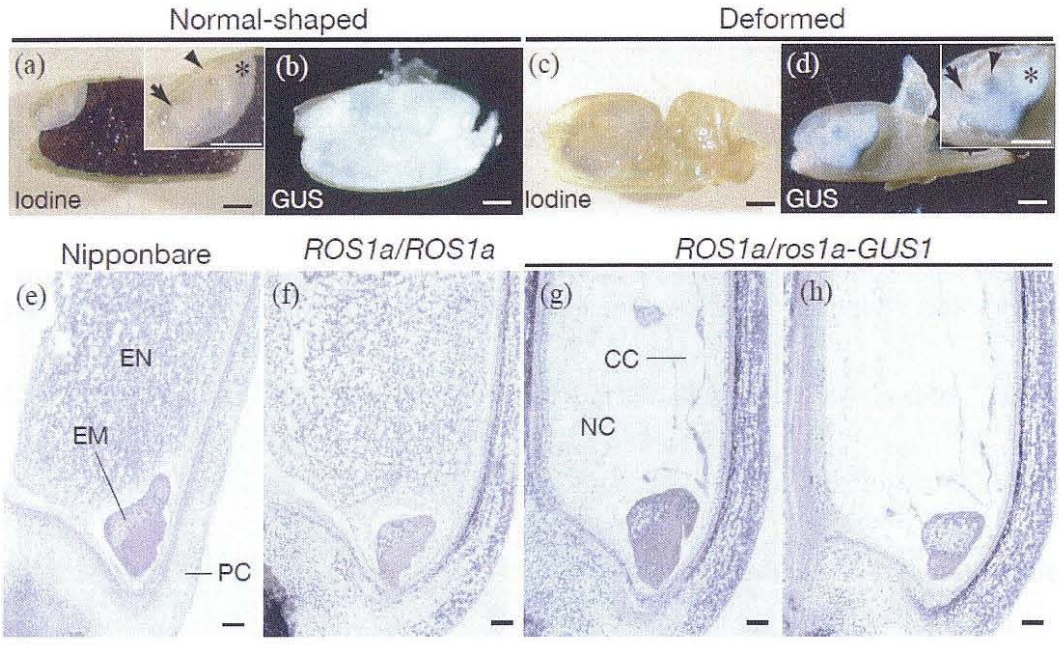

Normal-shaped
Deformed
(a)
(b)
(c)
(d)
Iodine
GUS
Iodine
GUS
Nipponbare
ROS1a/ROS1a
ROS1a/ros1a-GUS1
(e)
(f)
(g)
(h)
EN
EM
PC
CC
NC

Figure 11

Figure 11. Effects of *ros1a-GUS1* on seed development.

(a-d) Normal-shaped (a,b) and deformed (c,d) seed morphologies obtained from self-pollinated *ROS1a/ros1a-GUS1* plants. Seeds were stained with iodine (a,c) or for GUS expression (b,d). Insets in (a) and (d) are magnified views of embryos. Arrow, arrowhead, and asterisk indicate radicle, coleoptile, and scutellum, respectively.

(e-h) Longitudinal sections of 5 DAP Nipponbare (e) and self-pollinated *ROS1a/ros1a-GUS1* (f-h) seeds. Assignment of deformed seeds (g,h) was based on the lack of endosperm development at 5 DAP, when endosperm was developing in the normal-shaped seeds (f). Sections were immunostained with anti-GUS. CC, central cavity; EM, embryo; EN, endosperm; NC, nucellus; PC, pericarp. Bars = 1 mm in (a-d) and insets, and 100 μm in (e-h).

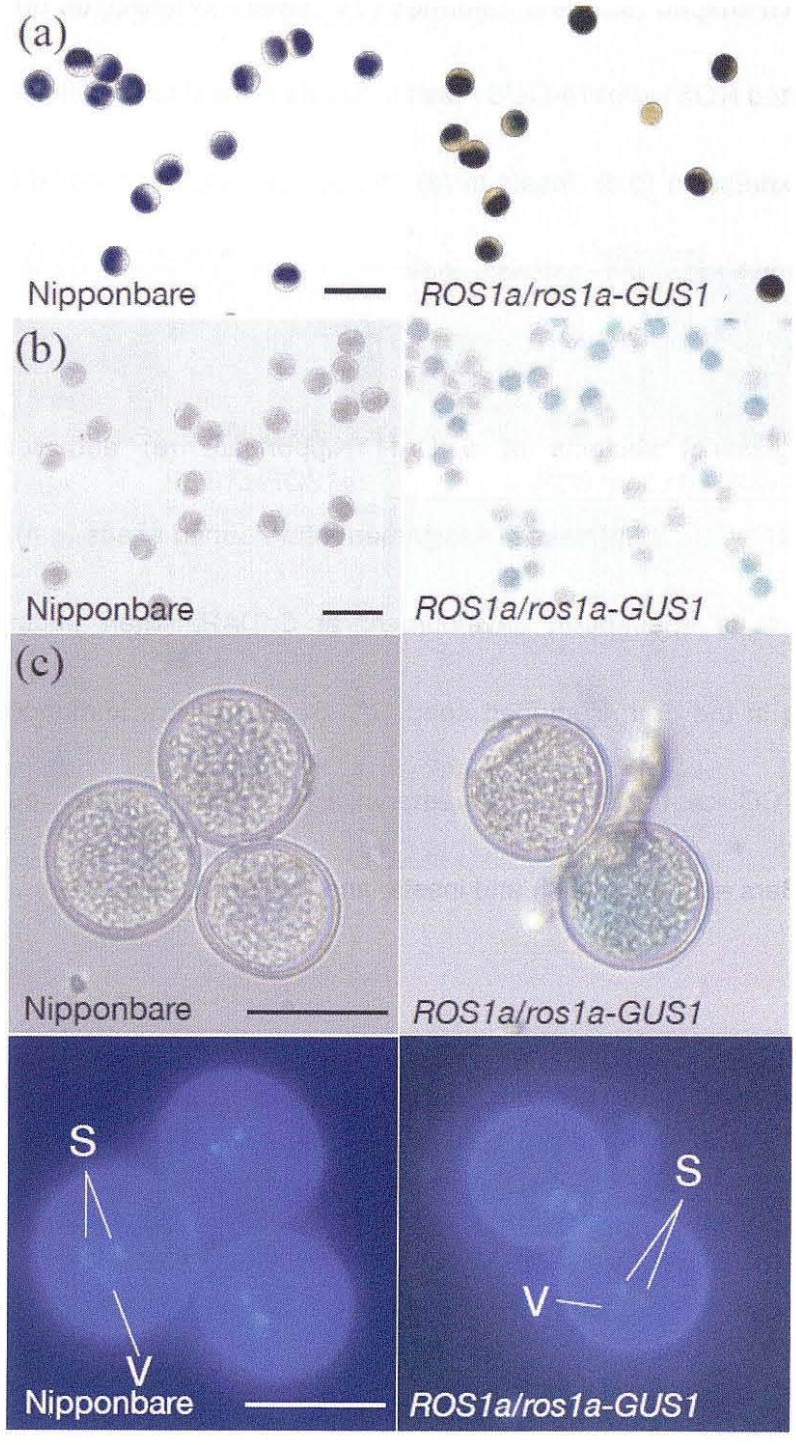

(a)
Nipponbare
ROS1a/ros1a-GUS1
(b)
Nipponbare
ROS1a/ros1a-GUS1
(c)
Nipponbare
ROS1a/ros1a-GUS1
S
V
Nipponbare
S
V
ROS1a/ros1a-GUS1

Figure 12

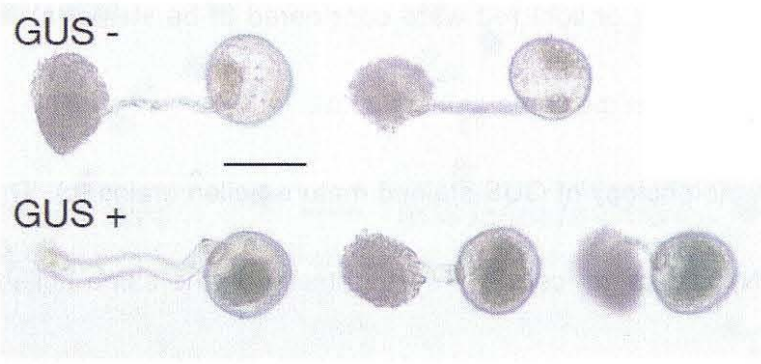
Figure 12. Effects of *ros1a-GUS1* on pollen viability and development.

(a,b) Mature pollen grains stained black with iodine were judged as viable, and those stained yellow or light red were considered to be sterile (a). GUS-stained pollen grains (b) from the same anthers in (a).

(c,d) Pollen morphology of GUS-stained mature pollen grains (c). The vegetative nucleus (VN) and sperm cells (S) were detected in the same pollen grains with DAPI (d). Bars = 100 μm in (a,b) and 50 μm in (c,d).

(a) GUS -

GUS +



(b) GUS+ (%)

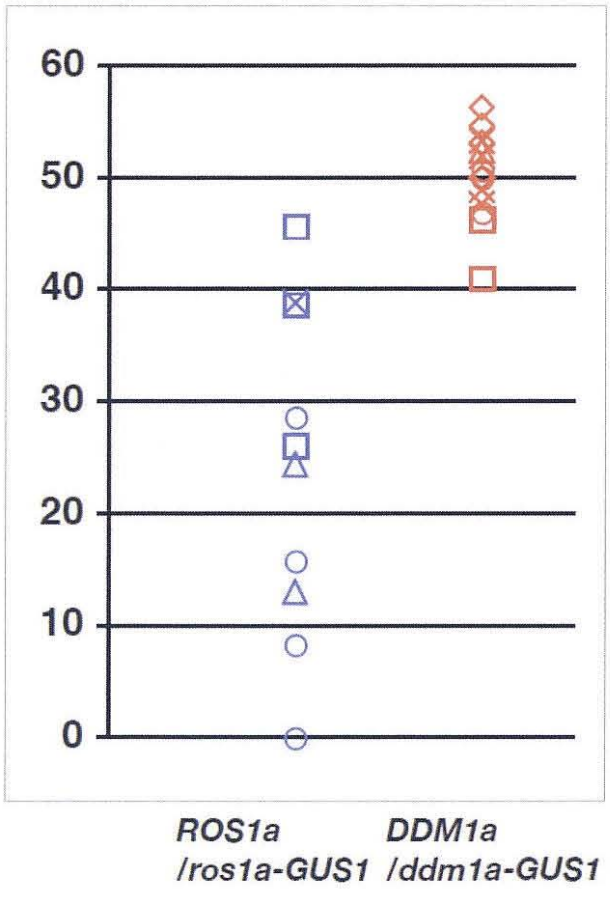


Figure 13

Figure 13. *In vitro* germination of pollen from *ROS1a/ros1a-GUS1* plants.

(a) Representative GUS-negative (GUS–) and GUS-positive (GUS+) germinated pollen.

(b) Ratios of GUS-positive to germinated pollen. Each plot indicates an independent pollen sample from an individual anther. The transgenic rice *DDM1a/ddm1a-GUS1* plants were obtained from Dr. Yasuyo Johzuka-Hisatomi. The same symbols represent pollen samples from the same plant lines. Bar = 50 μm in (a).

Table 1. Segregation analysis of T_1 seedlings derived from self-pollinated T_0 *ROS1a/ros1a-GUS1* plants by PCR genotyping

ROS1a/ros1a-GUS1 Selfed	Genotype			
Transgenic Plant Used	*ROS1a/ROS1a*	*ROS1a/ros1a-GUS1*	*ros1a-GUS1/ros1a-GUS1*	Total
L72-21	20	0	0	20
L83-2	20	0	0	20
L92-15	20	0	0	20
L95-1	5	1[a]	0	6
L95-2	26	1[a]	0	27
L95-4	12	0	0	12
L95-6	29	0	0	29
L95-23	38	0	0	38
L144-1	8	0	0	8
L144-12	17	0	0	17
L243-20	17	0	0	17
L243-28	18	0	0	18
Total	230	2	0	232

The presence of the *ROS1a* or *ros1a-GUS1* allele was examined by PCR genotyping (Figure 6a,c). [a]The *ros1a-GUS1* allele in these T_1 progeny was found to be not transmittable.

Table 2. Segregation of T_1 seeds derived from self-pollinated T_0 *ROS1a/ros1a-GUS1* plants

ROS1a/ros1a-GUS1 selfed		Phenotype		
Plant Used	Infertile	Normal-Shaped	Deformed	Total
N	4	38 (0)	0 (0)	42
L95-18	10	13 (0)	11 (11)	34
L144-12	6	15 (0)	19 (19)	40
L243-20	6	11 (0)	9 (9)	26
L243-28	5	15 (0)	17 (17)	37
Total	27	54 (0)	56 (56)	137

Typical examples of infertile flower remnants and normal-shaped and deformed seeds are shown in Figure 10, and see also Figure 11. All spikelets were examined, and numerals in parentheses indicate GUS-positive seeds. N, Nipponbare.

Table 3. Segregation of T_1 seeds in reciprocal crosses

		Phenotype		
ROS1a/ros1a-GUS1 (♀) x N (♂)	Infertile	Normal-Shaped	Deformed	Total
L144-12	17	6 (0)	12 (12)	35
L243-20	13	5 (0)	7 (7)	25
L243-22	15	9 (0)	15 (15)	39
L243-28	21	4 (0)	7 (7)	32
Total	66	24 (0)	41 (41)	131
N (♀) x *ROS1a/ros1a-GUS1* (♂)				
L144-12	10	30 (0)	0 (0)	40
L243-20	39	19 (0)	0 (0)	58
L243-22	32	15 (0)	0 (0)	47
L243-28	24	16 (0)	0 (0)	40
Total	105	80 (0)	0 (0)	185

The assignments of phenotypes are as in Table 2. The numerals in parentheses indicate the GUS-positive seeds. The rate of infertility in the wild-type outcrosses (Nipponbare x Nipponbare) is 31/56, and the higher infertility rate in the outcross, compared with that in self-pollination in Table 2, is probably due to the cross-pollination procedure, in which undesirable damages in pistils may occur during emasculation. N, Nipponbare.

Table 4. Pollen viability and GUS expression at mature stage in *ROS1a/ros1a-GUS1* plants

	Iodine				GUS			
	Broken	Yellow or Light Red	Black	Total	Broken	–	+	Total
N	2	2 (3.1%)	129 (96.9%)	133	6	333 (100%)	0 (0%)	339
L95-23	29	35 (15.3%)	357 (84.7%)	421	31	120 (49.6%)	122 (50.4%)	273
L144-19	1	5 (6.1%)	93 (93.9%)	99	19	199 (55.6%)	159 (44.4%)	377
L243-20	13	18 (6.1%)	479 (93.9%)	510	14	110 (47.2%)	123 (52.8%)	247
L243-23	5	10 (7.8%)	448 (96.7%)	463	32	305 (52.5%)	276 (47.5%)	613
Total	48	68 (7.8%)	1377 (92.2%)	1493	96	734 (51.9%)	680 (48.1%)	1510

Pollens obtained from anthers contained a few broken pollens, and the remaining unbroken pollens were subjected to iodine or GUS staining. Pollen viability was judged by iodine staining (Figure 12). Numbers in parentheses indicate the percentages of the pollens displaying the indicated traits in the unbroken pollen population. N, Nipponbare.

Table 5. List of primers

Designation	Sequences	Designation	Sequences	Comments
Primer sets for PCR and RT-PCR amplification				
5F PacI	TTAATTAAGGCCATGAGCTTAATCACTT	5R SrfI	GCCCGGGCATTTCCTTTCACTACTTCAT	5' homologous region of *ROS1a* in pJHY-Ki-ROS1a (Figure 6b)
C5F I-CeuI	TAACTATAACGGTCCTAAGGTAGCGAGA CAATATCATGAATACTAG	C5R PacI	TTAATTAACTTGGGCAAAATAAGGCTTTC	Additional 5' homologous region of *ROS1a* in pJHY-Ki-ROS1aC
3F PmeI	GTTTAAACATGCAGGATTTTGGACAATG	3R AscI	GGCGCGCCTCCTGCAAACATAGAGGTAA	3' homologous region of *ROS1a* in pJHY-Ki-ROS1a (Figure 6b)
C3F AscI	GGCGCGCCATCGGCGTTTTTCTCCTTGG	C3R I-SceI	ATTACCCTGTTATCCCTACATCTTCCGCT TCGACAGCG	Additional 3' homologous region of *ROS1a* in pJHY-Ki-ROS1aC
S5F	TTCAGGGAAGGGGTGAATTC	SGR	ATTCCACAGTTTTCGCGATC	5' junction fragment (Figure 6c)
S3F	GTATAATGTATGCTATACGAAGTTATGTTT	S3R	TGCCAATGTCTTCGGCAGTG	3' junction fragment (Figure 6c)
KIF	CACTGCATTATTGCTAGTCC	SGR	ATTCCACAGTTTTCGCGATC	Genotyping for the *ros1a-GUS1* allele (Figure 6c; Table 1)
WTF	ATGTACTTGAGGGAAGTTG	WTR	TCTGTTGAAGGTATTGGAG	Genotyping for the *ROS1a* allele (Figure 6a; Table 1)
dig5F	GGGCTGTTTGTCTTGTTTAC	dig5R	CTTTCACTACTTCATCTCTC	5' region probe (Figure 6c; Figure 7)
digGUSF	ATTGCTGTGCCAGGCAG	digGUSR	TAACCTTCACCCGGTTG	GUS probe (Figure 6c; Figure 7)
dighygF	CTCGTGCTTTCAGCTTCGATGTAGG	dighygR	GAAGATGTTGGCGACCTCGTATTGG	HPT probe (Figure 1c; Figure 7)
dig3F	ATGCAGGATTTTGGACAATG	dig3R	TGCCTGTTCTGCAGGAG	3' region probe (Figure 6c; Figure 7)
GF	CGCTCACACCGATACCATCA	GR	CGGCTGATGCAGTTTCTCCT	RT-PCR for *GUS* (Figure 10)
qRF	CCTAAGGACAGAGCACCAAG	qRR	GCACATAGTTCACCATTCTC	RT-PCR and qRT-PCR for *ROS1a* (Figures 1a, 1c, and 3a; Figure 4 and 8)
qROS1bF	CTATATGGACCCCAGGCAAG	qROS1bR	AGACTGCATTTCCCGTGTAC	qRT-PCR for *ROS1b* (Figure 4)
qROS1cF	TGCTGAATACCATGCACGAG	qROS1cR	TAAAGCTACGAGATTGCGGC	qRT-PCR for *ROS1c* (Figure 4)
qROS1dF	TGCAGGTTGAACATGGGCAG	qROS1dR	TCGTTGTGATCATGAATGTG	qRT-PCR for *ROS1d* (Figure 4)
qDML3aF	TGCATCTGGAACCTAGATAG	qDML3aR	ATTTTCACCTGTTCCGTTCC	qRT-PCR for *DML3a* (Figure 4)
qDML3bF	AAGACGATGAGAATGGGTAC	qDML3bR	CACCATTGGAGTTGATTTGG	qRT-PCR for *DML3b* (Figure 4)
act F	GAGTATGATGAGTCGGGTCCAG	act R	ACACCAACAATCCCAAACAGAG	qRT-PCR for *Actin* (Figure 4 and 7)
cD5FXmnI	GAAGGATTTCTATGCAGGATTTTGGACAATGG	cD3RSalI	GTCGACCTATTCATCATCTCTTCC	Outside primers for full-length cDNA of *ROS1a* (Figure 5)
cD5FXmnI	GAAGGATTTCTATGCAGGATTTTGGACAATGG	cDDAR	CTACCAACATTTGTAGCCACAGG	5' segment containing D1515A point mutation (Figure 5)
cDKAF	GACTTGGACTTGCAAGTGTTGAG	cD3RSalI	GTCGACCTATTCATCATCTCTTCC	3' segment containing K1499A point mutation (Figure 5)
Primers for 5'-RACE analysis				
5RA1	TGGTCTGAGATTGAGGCAGCCATTG			Primary amplification
5RA2	TTGAGGCAGCCATTGTCCAAAATCC			Nested amplification
Primers for 3'-RACE analysis				
3RA1	CACTGTATGCAAGACTCCACTTTCC			Primary amplification
3RA2	GCTCCAGGAAGAGATGATGAATAGG			Nested amplification

DISCUSSION

Using HR-promoted knock-in GT, whereby *GUS* was fused with the endogenous promoter of a target gene (Figure 6), an identical null mutation for the rice *ROS1a* gene encoding a putative cytosine DNA demethylase, whose activity was implicated by the toxic effects on *E. coli* containing 5-meC in their genome (Figure 5) were reproducibly generated. The obtained null mutation, *ros1a-GUS1*, virtually failed to be transmitted to progeny (Tables 1 and 2). The knock-in T_0 mutants (*ROS1a/ros1a-GUS1*) were subsequently crossed with Nipponbare (*ROS1a/ROS1a*), and the results clearly showed that both paternal and maternal *ros1a-GUS1* alleles were unable to be transmitted to progeny (Table 3), indicating that DNA demethylation must play important roles in both male and female gametophytic generation. Indeed, the GUS staining patterns of the knock-in T_0 plants indicated that *ROS1a* was expressed in pollen and unfertilized ovules as well as in meristematic cells (Figure 9).

Even in the presence of the wild-type paternal *ROS1a* allele, the maternal null *ros1a-GUS1* allele caused failure of early stage endosperm development (Figures 11c,d and 10d-n), indicating nonequivalent contribution of maternal and paternal *ROS1a* to endosperm development. Thus, *ros1a-GUS* in rice seems to exert comparable effects to that of *dme* in Arabidopsis in their female gametophytes. Both *ros1a-GUS* and *dme* cause maternal allele-specific endosperm defects:

ros1a-GUS causes severely underdeveloped endosperm (Figures 11 and 10), whereas *dme* causes an enlarged and then collapsed endosperm (Choi *et al.*, 2002). Although phylogenic analysis indicates that the rice genome contains no DME orthologs (Zemach *et al.*, 2010), their maternal allele-specific defects in endosperm development implies that *ROS1a* and *DME* may play certain analogous roles.

DME is preferentially expressed in the homodiploid central cell of the female gametophyte before fertilization and promotes maternal allele-specific global hypomethylation, which leads to the maternal allele-specific expression of imprinted genes, including the Polycomb Repressive Complex 2 (PRC2) genes containing *MEDEA* (*MEA*) and *FERTILIZATION INDEPENDENT SEED 2* (*FIS2*), and transposons in the endosperm (Choi *et al.*, 2002; Gehring *et al.*, 2006, 2009a; Jullien *et al.*, 2006). The expression of these imprinted genes is prerequisite for normal endosperm development. The *mea* or *fis2* mutant exhibited characteristic maternal defects in endosperm similar to the *dme* mutant, because PRC2 represses the autonomous growth of unfertilized endosperm and the overgrowth of fertilized endosperm (Luo *et al.*, 1999; Ohad *et al.*, 1999). For rice, GUS staining was detected in the central cells of *ROS1a/ros1a-GUS1* plants (Figure 9f,g), indicating that *ROS1a* is expressed in the central cell of the female gametophyte. Although many maternally expressed imprinted genes were

identified in the rice endosperm, very few of them shared in common with those in Arabidopsis (Luo *et al.*, 2011). The expression of *FERTILIZATION INDEPENDENT ENDOSPERM 1* (*FIE1*) (LOC_Os08g04290), one of such imprinted genes encoding a component of PRC2 in rice, appears to correlate with DNA demethylation at its 5' region, but an *fie1* disruptant conferred no defects in endosperm development (Luo *et al.*, 2009; Ishikawa *et al.*, 2011). The results implied that certain PRC2 roles of rice in endosperm development may differ from those of Arabidopsis. Although I could not demonstrate ROS1a-promoted DNA demethylation, *ROS1a* is postulated to activate the imprinted genes through DNA demethylation in order to facilitate normal endosperm development. GUS staining was also detected in the antipodals and egg apparatus comprising two synergids and an egg cell whose genetic information is transmittable to progeny (Figure 9f,g). The molecular mechanisms of ROS1a function in these cells remain to be elucidated.

Although *DME* in Arabidopsis is predominantly expressed in the central cell of the female gametophyte and is only marginally expressed to demethylate some imprinted genes and transposons in the vegetative cell of the male gametophyte (Choi *et al.*, 2002; Schoft *et al.*, 2011), comparable amounts of *ROS1a* were expressed in both the female and male gametophytes (Figures 1c and 9c,d), and the paternal *ros1a-GUS1* allele could not be transmitted to progeny (Tables 1 and

3). Although the molecular mechanisms for the inability to transmit the paternal *ros1a-GUS1* allele are unknown, the *ros1a-GUS1* pollen germinated less efficiently than did the *ROS1a* pollen. In certain Arabidopsis ecotypes, Col-*gl and* Col-0, reduced germination and transmission of *dme* pollen were also observed (Schoft *et al.*, 2011). The vegetative nucleus in pollen supports the sperm cell before fertilization and has reduced DNA methylation in Arabidopsis, resulting in transient reactivation of diverse transposons and the generation of small interfering RNAs that may ensure the genomic integrity of sperm cells by silencing transposons in the sperm cells (Slotkin *et al.*, 2009; Law and Jacobsen, 2010). Whether similar DNA demethylation occurs in the vegetative nucleus in rice pollen and whether the ROS1a proteins promote such demethylation remain to be determined. If such DNA demethylation occurs in the vegetative nucleus, it will be intriguing to ask whether such demethylation can directly correlate to the intransmittable nature of the paternal *ros1a-GUS1* allele (Tables 1 and 3).

qRT-PCR analysis revealed that *ROS1a, ROS1c, ROS1d,* and *DML3a* appear to be expressed in the same tissues examined, including anthers and pistils where *ROS1a* is the most abundantly expressed gene (Figure 4). The same analysis also showed that *ROS1c*, *ROS1d*, and *DML3a* were moderately expressed in both pistils and immature seeds two days after pollination. It was reported recently that the null mutants of *ROS1c* encoding a 5-meC DNA

glycosylase/lyase have no effects on transmission of their null alleles and produce a small portion of wrinkled seeds (La *et al.*, 2011). Their results are in sharp contrast to those in the null allele of *ROS1a* described here and indicate that the moderate expression of *ROS1c, ROS1d*, and *DML3a* in pistils could not compensate for the loss of *ROS1a*. It is interesting to ask whether both *ROS1a* and any of other three genes are expressed simultaneously in the same cells in the pistil. If *ROS1a* and *ROS1c* are expressed in the same cells, one would have to further ask whether the different transmissibility of their null alleles could be explained simply by their expression levels or by intrinsic differences in their molecular functions.

The null *ros1a-GUS1* allele was unable to be transmitted to progeny (Table 1). In conventional mutagenesis procedures, mutants are usually identified and isolated as segregants in the progeny population (e.g., Leung and An, 2004; Upadhyaya, 2007). It is clear that such procedures would not have isolated the *ros1a-GUS1* allele. In contrast, our HR-promoted GT procedure (Terada *et al.*, 2002, 2007; Johzuka-Hisatomi *et al.*, 2008; Yamauchi *et al.*, 2009) allowed T_0 transgenic plants to be reproducibly obtained with the exact designed structure of the targeted gene in the heterozygous condition. Moreover, the knock-in targeting strategy allowed detection of spatiotemporal *ROS1a* expression and tracking of the transmitted *ros1a* allele. Because GUS staining was necessary for detecting

ROS1a expression, it was, however, difficult to analyze *in vivo* and/or to isolate viable tissues that express *GUS* for detailed examination. Utilization of other markers such as green fluorescent protein, in combination with other techniques including laser micro dissection and/or fluorescence activated cell sorting, would facilitate further characterization of *ROS1a* function in endosperm development and pollen gametophytic transmission.

REFERENCES

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

Chan, S.W.-L., Henderson, I.R. and Jacobsen, S.E. (2005) Gardening the genome: DNA methylation in *Arabidopsis thaliana*. *Nat. Rev. Genet.* **6**, 351–360.

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

Choi, Y., Harada, J.J., Goldberg, R.B. and Fischer, R.L. (2004) 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**, 7481–7486.

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

Gehring, M., Bubb, K.L. and Henikoff, S. (2009a) Extensive demethylation of

repetitive elements during seed development underlies gene imprinting. *Science*, **324**, 1447-1451.

Gehring, M., Reik, W. and Henikoff, S. (2009b) DNA demethylation by DNA repair. *Trends Genet.* **25**, 82-90.

Gong, Z., Morales-Ruiz, T., Ariza, R.R., Roldán-Arjona, T., David, L. and Zhu, J.-K. (2002) *ROS1*, a repressor of transcriptional gene silencing in *Arabidopsis*, encodes a DNA glycosylase/lyase. *Cell*, **111**, 803-814.

Henderson, I.R. and Jacobsen, S.E. (2007) Epigenetic inheritance in plants. *Nature*, **447**, 418-424.

Hsieh, T.-F., Ibarra, C.A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R.L. and Zilberman, D. (2009) Genome-wide demethylation of *Arabidopsis* endosperm. *Science*, **324**, 1451-1454.

Ishikawa, R., Ohnishi, T., Kinoshita, Y., Eiguchi, M., Kurata, N. and Kinoshita, T. (2011) Rice interspecies hybrids show precocious or delayed developmental transitions in the endosperm without change to the rate of syncytial nuclear division. *Plant J.* **65** 798-806.

Johzuka-Hisatomi, Y., Terada, R. and Iida, S. (2008) Efficient transfer of base changes from a vector to the rice genome by homologous recombination: involvement of heteroduplex formation and mismatch correction. *Nucleic Acids Res.* **36**, 4727-4735.

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

La, H., Ding, B., Mishra, G.P., *et al.* (2011) A 5-methylacytosine DNA glycosylase/lyase demethylates the retrotransposon *TOS17* and promotes its transposition in rice. *Proc. Natl. Acad. Sci. USA*, **108**, 15498-15503.

Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204-220.

Leung, H. and An, G. (2004) Rice functional genomics: large-scale gene discovery and applications to crop improvement. *Adv. Agron.* **82,** 55-111.

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

Luo, M., Platten, D., Chaudhury, A., Peacock, W.J. and Dennis, E.S. (2009) Expression, imprinting, and evolution of rice homologs of the polycomb group genes. *Mol. Plant*, **2**, 711-723.

Luo, M., Taylor, J.M., Spriggs, A., Zhang, H., Wu, X., Russel, S., Singh, M and Koltunow, A. (2011) A genome-wide survey of imprinted genes in rice

seeds reveal imprinting primarily occurs in endosperm. *PLoS Genet.* **7**, e1002125. doi: 10.1371/journal.pgen.1002125

Martienssen, R.A. and Colot, V. (2001) DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science*, **293**, 1070-1074.

Morales-Ruiz, T., Ortega-Galisteo, A.P., Ponferrada-Marín, M.I., Martínez-Macías, M.I., Ariza, R.R. and Roldán-Arjona, T. (2006) *DEMETER* and *REPRESSOR OF SILENCING 1* encode 5-methylcytosine DNA glycosylases. *Proc. Natl. Acad. Sci. USA*, **103**, 6853-6858.

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J.J., Goldberg, R.B., and Fischer, R.L. (1999) Mutations in *FIE*, a WD *Polycomb* group gene, allow endosperm development without fertilization. *Plant Cell* **11**, 407-416.

Ortega-Galisteo, A.P., Morales-Ruiz, T., Ariza, R.R. and Roldán-Arjona, T. (2008) Arabidopsis DEMETER-LIKE proteins DML2 and DML3 are required for appropriate distribution of DNA methylation marks. *Plant Mol. Biol.* **67**, 671–681.

Palmer, B.R. and Marinus, M.G. (1994) The *dam* and *dcm* strains of *Escherichia coli* – a review. *Gene* **143**, 1-12.

Penterman, J., Zilberman, D., Huh, J.H., Ballinger, T., Henikoff, S. and Fischer, R.L. (2007) DNA demethylation in the *Arabidopsis* genome. *Proc.*

Natl. Acad. Sci. USA, **104**, 6752-6757.

Schoft, V.K., Chumak, N., Choi, Y., *et al.* (2011) Function of DEMETER DNA glycosylase in the *Arabidopsis thaliana* male gametophyte. *Proc. Natl. Acad. Sci. USA*, **108**, 8042-8047.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzic, M., Becker, J.D., Feijó, J.A. and Martienssen, R.A. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**, 461-472.

Stemmer, W.P.C., Crameri, A., Ha, K.D., Brennan, T.M. and Heyneker, H.L. (1995) Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* **164,** 49-53.

Terada, R., Urawa, H., Inagaki, Y., Tsugane, K. and Iida, S. (2002) Efficient gene targeting by homologous recombination in rice. *Nat. Biotechnol.* **20**, 1030-1034.

Terada, R., Johzuka-Hisatomi, Y., Saitoh, M., Asao, H. and Iida, S. (2007) Gene targeting by homologous recombination as a biotechnological tool for rice functional genomics. *Plant Physiol.* **144**, 846-856.

Upadhyaya, N.M. (ed.) (2007) *Rice Functional Genomics - Challenges, Progress and Prospects.* New York: Springer.

Yamauchi, T., Johzuka-Hisatomi, Y., Fukada-Tanaka, S., Terada, R., Nakamura, I. and Iida, S. (2009) Homologous recombination-mediated

knock-in targeting of the *MET1a* gene for a maintenance DNA methyltransferase reproducibly reveals dosage-dependent spatiotemporal gene expression in rice. *Plant J.* **60**, 386-396.

Zemach, A., Kim, M.Y., Silva, P., Rodrigues, J.A., Dotson, B., Brooks, M.D. and Zilberman, D. (2010) Local DNA hypomethylation activates genes in rice endosperm. *Proc. Natl Acad. Sci. USA*, **107**, 18729-18734.

Zhu, J.-K. (2009) Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43,** 143-166.

Chapter 2

DDM1a and DDM1b, chromatin remodeling factors, modulate cytosine methylation patterns and transcriptional reactivation at repeat sequences in rice

INTRODUCTION

Cytosine methylation is a DNA base modification with roles in variety of biological processes such as development, disease, and silencing transposable elements (TEs) and repetitive sequences in plants as well as in animals and fungi (Law and Jacobsen, 2010). In plants, CG methylation is commonly found within gene bodies, whereas non-CG methylation, CHG and CHH (where H is A, C, or T), is enriched in TEs and repetitive sequences (Law and Jacobsen, 2010). The genes and their corresponding encoded enzymes that mediate DNA methylation and demethylation have been characterized mainly in Arabidopsis, whose genome has fewer repetitive sequences than in any other plant species (Chan *et al.*, 2005; Zhu, 2009; Law and Jacobsen, 2010). The *DDM1* (*DECREASE IN DNA METHYLATION 1*) gene of Arabidopsis, which encodes a SWI2/SNF2-like chromatin remodeling ATPase, was identified by the isolation of DNA hypomethylation mutants (Vongs *et al.*, 1993; Jeddeloh *et al.*, 1999; Brzeski and Jerzmanowski, 2003). The *DDM1* is the major master regulator of TE activity in Arabidopsis, as most TEs lose DNA methylation, repressive histone modifications, and 24nt small interfering RNAs (siRNAs) in *ddm1* mutants (Gendrel *et al.*, 2002; Lippman *et al.*, 2003, 2004). The *ddm1* mutation also affects on single copy sequences after repeated self-pollination (Vongs *et al.*, 1993; Kakutani *et al.*, 1996). Decreased expression of *DDM1* orthologs of *Brassica rapa*, *BrDDM1a* and

BrDDM1b, induces hypomethylation in TEs and repetitive sequences, and transcriptional reactivation of some TEs (Fujimoto *et al.*, 2008). Mutation of the mouse *Lymphocyte-specific helicase* (*Lsh*) gene (an ortholog of *DDM1*) causes a substantial demethylation of the genome. Lsh is involved in mediating the epigenetic silencing of tandem repeats at centromeric heterochromatin as well as TEs through its primary role in the establishment of DNA methylation patterns at repetitive elements in mammalian oocytes (Dennis *et al.*, 2001; Muegge, 2005; De la Fuente *et al.*, 2006). These results suggest that mechanisms controlling DNA methylation are conserved, and drawing on insights from not only Arabidopsis but also from different plant species or animals should deepen our understanding of the regulation and biological significance of DNA methylation.

In chapter 1, I described homologous recombination (HR)-promoted gene targeting (GT) with positive-negative selection in rice. Compared with Arabidopsis, rice has bigger genome size and more repetitive sequences; over 40 % of its genome is repetitive DNA and most of this is related to TEs (Goff *et al.*, 2002; Yu *et al.*, 2002). In this study, to reveal the effect of OsDDM1 on the rice genome, knock-out targeting to obtain a mutant that disrupts each of two *OsDDM1* genes, *OsDDM1a* and *OsDDM1b,* was employed. Primary (T_0) transgenic plants with knock-out allele, *Osddm1a* or *Osddm1b*, in the heterozygous condition were reproducibly obtained. Subsequently I obtained

disruptant mutants for single genes *Osddm1a*, *Osddm1b*, and double mutant *Osddm1a Osddm1b*. Analysis of these disruptant mutants indicated an essential role for OsDDM1s on epigenetic regulation of repetitive sequences with different degree of contribution and functional redundancy between OsDDM1a and OsDDM1b.

MATERIALS AND METHODS

Nucleic acid procedures

General nucleic acid procedures, including plasmid preparation, plant DNA and RNA preparation, PCR and RT-PCR amplification, and Southern blot and DNA sequencing analyses were performed as described before (Terada *et al.*, 2007; Johzuka-Hisatomi *et al.*, 2008). The sequences of primers are listed (Table 1). Multiple amino-acid alignments were produced with a web-based version of ClustalW (http:/clustalw.ddbj.nig.ac.jp/) using default settings.

Knock-out targeting

Vector construction

The vector used here for targeting of *OsDDM1a*, pJHYDDM1a, is a derivative of the backbone of the targeting vector pINA134 (Terada *et al.*, 2002, 2007), which carries a 3.0-kb *Hind*III-*Srf*I fragment of the *OsDDM1a* promoter and a 3.0-kb *Pme*I-*Asc*I fragment of the 5' coding region of *OsDDM1a* (Figure 1b). The 3.0-kb fragments were prepared by PCR amplification of the Nipponbare genomic sequence with appropriate primers including the indicated restriction sites (Table 1) used for inserting into the corresponding cloning sites of pINA134. The control plasmid pJHYDDM1aC, a pJHYDDM1a derivative that includes a 3.5-kb fragment from the 5' *OsDDM1a* coding region to detect authentic 3'-JF (Figure 1c), was

constructed in the same way by cloning the 3'-flanking 1.0-kb *Asc*I-I-*Sce*I fragments of the inserted 3.0-kb *Pme*I-*Asc*I fragments. The vector used here for targeting of *OsDDM1b*, pJHY-Ki-DDM1b, is a derivative of the backbone of the targeting vector pJHY-Ki2 (Yamauchi *et al.*, 2009), which carries a 3.0-kb *Pac*I-*Hind*III fragment of the *OsDDM1b* promoter and a 3.0-kb *Pme*I-*Asc*I fragment of the 5' coding region of *OsDDM1b* (Figure 1e). The 3.0-kb fragments were prepared by PCR amplification of the Nipponbare genomic sequence with appropriate primers including the indicated restriction sites (Table 1) used for inserting into the corresponding cloning sites of pJHY-Ki2. Sequencing of the cloned fragments showed no base changes. The control plasmid pJHY-Ki-DDM1bC, a pJHY-Ki-DDM1b derivative, that includes a 3.5-kb *OsDDM1b* promoter fragment and a 3.5-kb fragment from the 5' *OsDDM1b* coding region to detect authentic 5'-JF and 3'-JF (Figure 1f), was constructed in the same way by cloning the 5'-flanking 0.5-kb I-*Ceu*I-*Pac*I and 3'-flanking 0.5-kb *Asc*I-I-*Sce*I fragments of the inserted 3.0-kb *Pac*I-*Srf*I and 3.0-kb *Pme*I-*Asc*I fragments, respectively.

Plant transformation

Agrobacterium-mediated rice transformation and screening for targeted calli were performed and characterized as described (Terada *et al.*, 2007; Yamauchi *et al.*,

2009). Targeted calli were isolated by PCR analysis detecting 3.46-kb 5'- and 4.09-kb 3'-JFs for *OsDDM1a*, and 3.5-kb 5'- and 3.5-kb 3'-JFs for *OsDDM1b* (Figure 1c,f) out of survived calli through positive-negative selection. From a targeted callus, several transgenic plants were regenerated through multiple shoots. These plants were subjected to Southern blot analysis to confirm that the regenerants contained expected knock-out allele with no additional undesirable ectopic events. The probes for Southern blot analysis were prepared by PCR amplification on pJHYDDM1aC or pJHY-KiDDM1bC. The genotype of *OsDDM1a* and/or *OsDDM1b* alleles were examined by the appearance of the 1- and 4.5-kb fragments for the *OsDDM1a* and *Osddm1a* knock-out alleles (Figure 1a,c), and 0.7- and 3.5-kb fragments for the *OsDDM1b* and *Osddm1b* knock-out alleles (Figure 1d,f), respectively with PCR amplification. Rice plants were grown at 25°C for 10 h in the dark, and at 25°C for 2 h and 30°C for 12 h in the light condition.

Genetic crosses

Osddm1a Osddm1b double mutant plants were generated as follows; F_1 progenies with the genotype *OsDDM1a/Osddm1a OsDDM1b/Osddm1b* by crossing T_0 *OsDDM1b/Osddm1b* plants to *Osddm1a* mutant plants in T2 generation were selected and self-fertilized. Resulting F_2 progeny plants were genotyped and used for DNA methylation assay. F_2 plants homozygous for both

Osddm1a and *Osddm1b* were then self-fertilized again and F_3 seedlings were used for microarray analysis.

DNA methylation assays

Genomic DNA was isolated from mature leaves emerged just before flag-leaves were emerged. Five μg aliquots of DNA were digested with *Hpa*II, *Hae*III or *Msp*I and electrophorested on 1.0 % agarose gels. The electrophoresed DNA was transferred onto Hybond-N^+ nylon membrane (GE Healthcare, Piscatway, NJ, USA), and hybridized with a digoxygenin-labeled probe at 65°C. After hybridization, the membrane was washed twice in a solution of 0.1 % SSC containing 0.1 % SDS at 65°C for 20 min. For hybridization probes, 5S rDNA, 45S rDNA, CentO and *RIRE7* fragments were PCR amplified from genomic DNA, and labeled with digoxygenin. The primers information for PCR amplification is shown in Table 1.

Microarray analysis

Microarray analysis was performed as described previously (Rajhi *et al.*, 2011).

RESULTS

Characterization of *OsDDM1*

In contrast to the single gene of *DDM1* in Arabidopsis, the available data indicate that the rice genome carries two *DDM1*-homologous genes: *OsDDM1a* (AB177380) located on chromosome 9 and *OsDDM1b* (AB177381) located on chromosome 3. The exon-intron structures are similar to each other; both *OsDDM1a* and *OsDDM1b* comprise 15 exons and 14 introns, and the nucleotide sequence identity of these two genes is 81 % including introns (Figure 1a,d). The predicted translation products from cDNA of *OsDDM1a* (AB177378) and *OsDDM1b* (AB177379) are 845 and 849 amino acids, respectively. The encoding amino acid identity between *OsDDM1a* and *OsDDM1b* is 93.9 % and those between *OsDDM1a* and Arabidopsis *DDM1* and between *OsDDM1b* and Arabidopsis *DDM1* were 69.3 and 68.4 %, respectively (Figure 2). Both *OsDDM1a* and *OsDDM1b* are expressed in all tissues examined (Figure 3). In vegetative tissues, both *OsDDM1a* and *OsDDM1b* transcripts are more abundant in young leaves than in older leaves. In reproductive tissues, transcripts become less abundant in immature seeds after pollination. The manner of transcription is similar between *OsDDM1a* and *OsDDM1b* during development with more abundant *OsDDM1b* transcript than *OsDDM1a*. Only exception is in anther where

OsDDM1a transcripts are considerably more abundant than those of *OsDDM1b* (Figure 3).

Generation of *OsDDM1s* targeted plants

In order to address the functions of OsDDM1a and OsDDM1b, we disrupted each of *OsDDM1* genes by knock-out GT, following the experimental design for previous studies (Terada *et al.*, 2007; Yamauchi *et al.*, 2009). To identify the targeted double crossovers by two consecutive homologous recombination events at the 3.0-kb homologous segments, the junction fragments 5'- JF or both 5'- JF and 3'-JF were PCR amplified (Figure 1c,f). Then 5'-JF and 3'-JF were sequenced at both ends to confirm that the PCR-amplified fragments were the expected junction fragments. By identifying the calli producing the junction fragments, we could get five *OsDDM1a* and 19 for *OsDDM1b* independent targeted calli with frequencies of 2.3×10^{-2} and 3.4×10^{-2} per surviving callus with positive-negative selection respectively. Four for *OsDDM1a* and 12 for *OsDDM1b* independently targeted T_0 plant lines were subsequently isolated from the targeted calli. Southern blot analysis of genomic DNAs from these T_0 plants revealed that they are with only one copy of *htp* disrupted the targeted gene in the heterozygous condition (*OsDDM1a/Osddm1a* or *OsDDM1b/Osddm1b*) through homologous recombination (Figure 1c,f and 4). No bands other than the

corresponding to the wild-type (WT) and knock-out allele were detectable, indicating that no additional ectopic events, such as a random integration of an additional copy of the transgenes, took place. Although *OsDDM1b* targeting construct was designed to produce transcriptional fusion to *GUS* gene, GUS activity was not detected in any tissues investigated. T_0 plants were fertile, and T_2 progenies were obtained by self-pollination of T_1 segregants. Disruptant mutants, *Osddm1a* (*Osddm1a/Osddm1a*) and *Osddm1b* (*Osddm1b/Osddm1b*), were also fertile in T_1 and T_2 generations. Compared with the sibling WT plants (*OsDDM1a/OsDDM1a* or *OsDDM1b/OsDDM1b*), no apparent phenotypic alteration could be detected in the *Osddm1a* or *Osddm1b* mutants in T_1 and T_2 generations.

Different contribution of OsDDM1a and OsDDM1b on DNA methylation

The decrease of cytosine methylation in the mature leaves of the disruptant mutant plants was examined by Southern blot analysis. Genomic DNA was extracted from mature leaves and digested by methylation sensitive restriction endonuclease *Hpa*II, which recognize the sequence, CCGG, and disables to cleave the site methylated at the inner or outer cytosine. The promoter region of 45S rDNA, the coding region of 5S rDNA and Centromeric tandem repeats O (CentO) were used as probes because these repeat regions are known to be

highly methylated in rice genome. The methylation levels of 5S rDNA and CentO in *Osddm1a* mutant plants in T_2 generation were comparable to these in WT and Nipponbare plants (Figure 5a,b). The 45S rDNA promoter is less methylated in all transgenic plants than in Nipponbare. Degree of hypomethylation in this region depends on plant lines rather than *OsDDM1a* targeted alleles (Figure 5c). I could detect the decrease of methylation level in this region in *alcohol dehydrogenase2* (*Adh2*) gene targeted plants (data not shown). Because it is known that Adh2 does not affect on 45S rDNA methylation, the decrease of DNA methylation levels in 45S rDNA might be due to the transformation events.

In *Osddm1b* mutant plants, in contrast with *Osddm1a* mutant plants, the methyaltion levels of 5S rDNA and CentO were considerably reduced compared with WT plants even in T_1 generation, whereas the methyaltion levels in WT plants were comparable to Nipponbare plants (Figure 6a,b). Although in Arabidopsis *ddm1* mutant, repeated self-pollination results in progressive demethylation in repetitive sequences, considerable progressive demethylation could not be detected in *Osddm1b* mutant plants at least between T_1 and T_2 generations (Figure 6a,b compare T_1 and T_2). In 45S rDNA, pararell result to *Osddm1a* mutant plants was observed in *Osddm1b* mutant plants. To see whether the methylation state of TEs is also affected in *Osddm1b* mutant plants, Southern blot analysis of retrotransposon *RIRE7* was performed, detecting the cytosine methylation in the

long terminal repeat (LTR) and *pol* regions (Figure 7). *RIRE7* is a component of the pericentromeric heterochromatin of rice chromosome (Kumekawa *et al.*, 2001) and the LTR is considered to be important for its transcription (Pouteau *et al.*, 1991; Takeda *et al.*, 1999). Compared with WT and Nipponbare plants, considerable hypomethylation in *Osddm1b* mutant plants was detected both in T_1 and T_2 generations (Figure 7b). In addition, I investigated transcriptional reactivation of several TEs including *RIRE7* by RT-PCR (Figure 8). Although the reduction of methylation in LTRs of *RIRE7* could be detected, considerable transcriptional reactivation of any TEs investigated in T_2 *Osddm1b* mutant plants was not detected.

Additive effects of OsDDM1a and OsDDM1b disruption on DNA methylation

Next, *Osddm1a Osddm1b* double disrupted mutant plants (*Osddm1a/Osddm1a*, *Osddm1b/Osddm1b*) were generated. By crossing T_0 *OsDDM1b/Osddm1b* plants to T_2 *Osddm1a* mutant plants, F_1 progenies with the genotype *OsDDM1a/Osddm1a OsDDM1b/Osddm1b* were obtained. By self-pollination of these plants, I could generate sibling plants with all different combinations of *OsDDM1a* and *OsDDM1b* alleles in the next generation from single parent plant. I performed Southern blot analysis of these selfed progenies. All 45S rDNA, 5S rDNA and CentO repeat regions were most hypomethylated in *Osddm1a*

Osddm1b double mutant plants compared with the any other sibling plants with different combinations of *OsDDM1a* and *OsDDM1b* alleles (Figure 9 and 10). Highly demethylation in *Osddm1a Osddm1b* double mutant plants was also detected in *RIRE7* (Figure 7). These results indicate that OsDDM1a and OsDDM1b work redundantly on all the investigated repeat regions. Notably, CentO sequence is more sensitive to the dose of *OsDDM1a* and *OsDDM1b* knock-out alleles than other repeat regions (Figure 9), although it is reported that *ddm1* allele is completely recessive in Arabidopsis.

Southern blot analysis of *Osddm1a Osddm1b* double mutant plants with additional methylation-sensitive restriction endonucleases, *Msp*I and *Hae*III was performed (Figure 11). *Msp*I is an isoschizomer of *Hpa*II. It recognizes CCGG and is disable to cleave the site methylated at the outer cytosine. *Hae*III recognizes GGCC and is disable to cleave if the inner cytosine is methylated. Digestion of CentO repeat region with the *Hae*III reports on methylation of CNN (in Nipponbare, the some *Hae*III site in CentO repeat is a CNN site). The methylation at *Msp*I site of CentO repeat region is decreased in *Osddm1a Osddm1b* double mutant plants relative to WT and Nipponbare plants. On the other hand, *Hae*III digestion of CentO repeats in double mutant plants occurred to the same extent as in WT and Nipponbare plants (Figure 11). The results indicate that OsDDM1s affect cytosine

methylation preferentially in CG and CNG sequence contexts, but do not affect or affect in small extent, if they do, in CNN sequence context.

Deregulation of transcripts in *Osddm1a Osddm1b* double mutant plants

Because *Osddm1a Osddm1b* double mutant plants could generate seeds, transcriptional reactivation of several TEs in *Osddm1a Osddm1b* double mutant plants in the next generation was investigated (Figure 8). Total RNA derived from *Osddm1a Osddm1b* mutant and Nipponbare seedlings was subjected to RT-PCR analysis. Transcriptional reactivations of a subset of TEs were detected in double mutant plants. Compared with Nipponbare plants, significantly high expression of *RIRE7* and *MuDR*-like (LOC_Os12g08080) was detected in double mutant plants, while the expression levels of *Tos17* in double mutant plants were comparable to these of in Nipponbare plants. Transcription of *iDart* could not be detected either in double mutant or Nipponbare plants. Taken together with the transcriptional reactivation in *Osddm1b* mutant plants (Figure 8), OsDDM1a and OsDDM1b affect on deregulation of TEs transcripts additively, and it is comparable to the additive effects on DNA methylation (Figure 7). The transcriptional change corresponds to the alterations in cytosine methylation.

In order to carry out a genome-wide search for other genes whose expression is affected in double mutant plants, microarray analysis was

performed and the gene expression patterns in double mutant was compared with Nipponbare plants (Table 2 and 3). Majority of sequences that were deregulated in mutant plants were increased in the disruption of *OsDDM1a* and *OsDDM1b*, suggesting that OsDDM1s mainly silences specific transcripts. Among the upregulated genes in double mutant plants, several TEs were detected (Table 2). In addition to TEs, many non-TE genes were also upregulated (Table 2). In cases of a subset of tissue-specific genes, the tissue-specific barrier of gene expression was released upon the disruption of both *OsDDM1a* and *OsDDM1b.* Thus, OsDDM1s deficiency appears to alter gene expression levels at repetitive sites and it influences gene expression pattern or/and levels of genes located near repetitive sites or/and downstream of *DDM1a DDM1b* genes. These results suggest that OsDDM1s play specific roles in protecting the chromatin state and function at repetitive elements in rice genome, which has many repetitive sequences, and, furthermore, protecting its genome environment from abberant gene expression caused by neighboring repetitive elements.

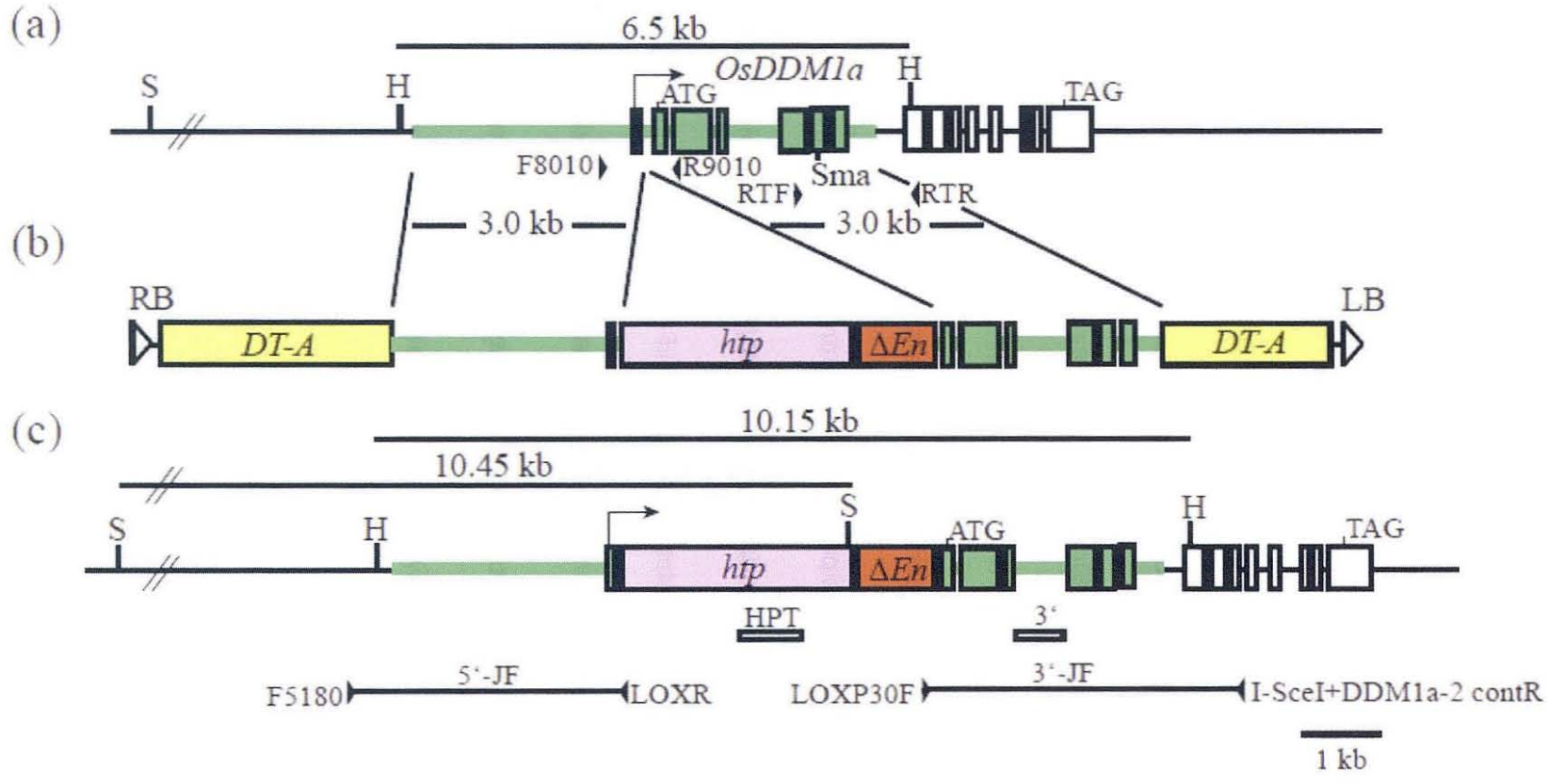

(a)
6.5 kb
S
H
OsDDM1a
H
ATG
TAG
F8010
R9010
Sma
RTF
RTR
(b)
3.0 kb
3.0 kb
RB
LB
DT-A
htp
ΔEn
DT-A
(c)
10.15 kb
10.45 kb
S
H
S
ATG
H
TAG
htp
ΔEn
HPT
3'
F5180
5'-JF
LOXR
LOXP30F
3'-JF
I-SceI+DDM1a-2 contR
1 kb

Figure 1

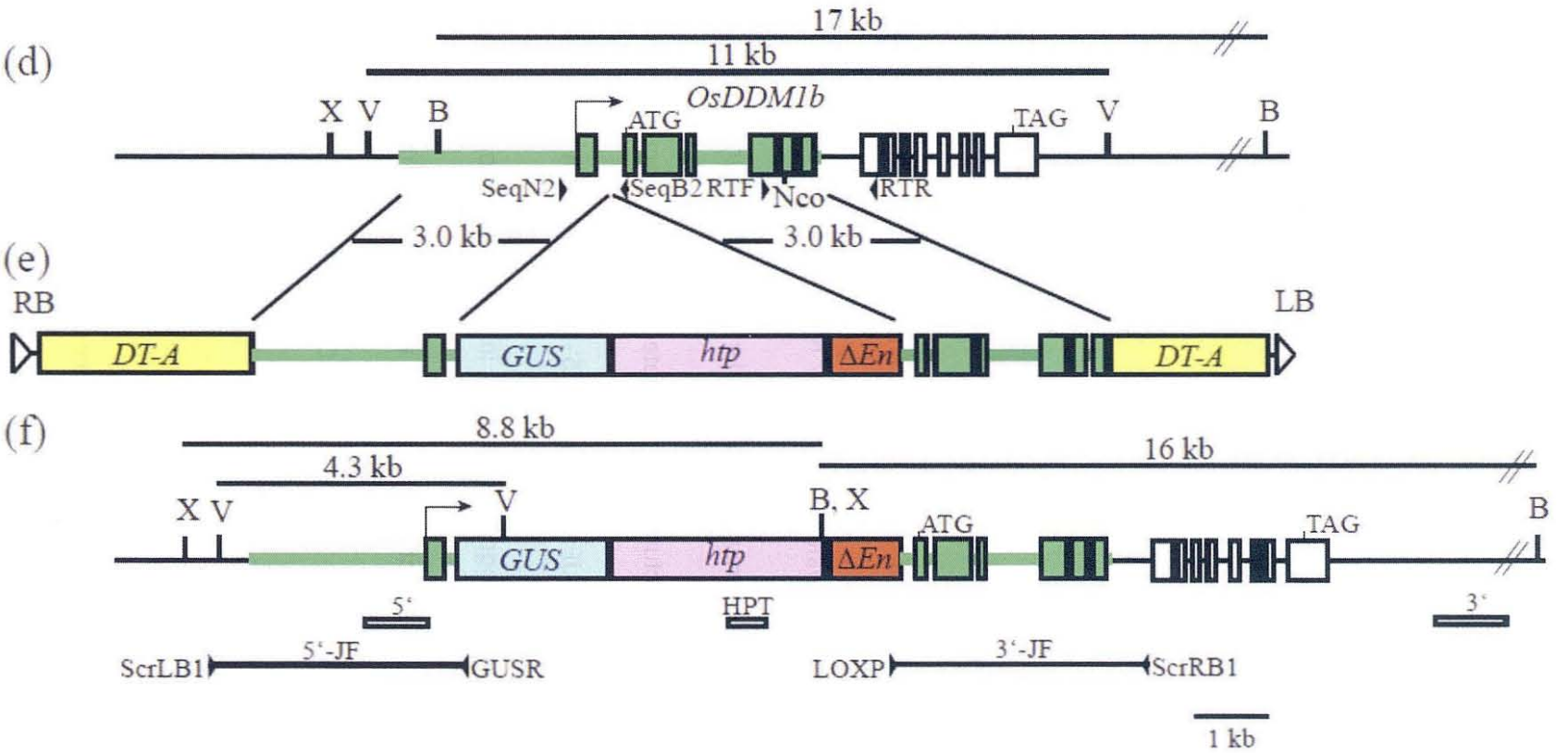

(d)
17 kb
11 kb
OsDDM1b
X V B
ATG
TAG
V
B
SeqN2
SeqB2
RTF
Nco
RTR
(e)
3.0 kb
3.0 kb
RB
LB
DT-A
GUS
htp
ΔEn
DT-A
(f)
8.8 kb
16 kb
4.3 kb
X V
V
B, X
ATG
TAG
B
GUS
htp
ΔEn
5'
HPT
3'
ScrLB1
5'-JF
GUSR
LOXP
3'-JF
ScrRB1
1 kb

Figure 1

Figure 1. Strategies for the gene targeting modification of the *OsDDM1a* (a-c) and *OsDDM1b* (d-f) in rice.

(a) and (d) Genomic structures of *OsDDM1a* (a) and *OsDDM1b* (d). Boxes and horizontal arrow indicate exons and transcriptional initiation, respectively. Green segments are also carried by targeting vector pJHYDDM1a (b) or pJHY-Ko-DDM1b (e). Arrowheads with RTF and RTR indicate primers for RT-PCR analysis.

(b) and (e) T-DNA regions of pJHYDDM1a (b) and pJHY-Ko-DDM1b (e). RB, right border; *DT-A*, *diphtheria toxin A* fragment gene for negative selection; *hpt*, *hygromycin phosphotransferase* gene for hygromycin B resistance; *ΔEn*, the 3' end of the maize *En* element; LB, left border; *GUS*, *β-glucuronidase* gene.

(c) and (f) Genomic structures of the targeted *OsDDM1a* locus (c) and *OsDDM1b* locus (f). The 5'- and 3'-JFs indicate the 5'- and 3'-junction fragments generated by HR between homologous segments in the endogenous target genes and targeting vectors, respectively. Flanking arrowheads of the F5180/LOXR and LOXP30F/I-SceI+DDM1a-2 contR for *OsDDM1a* and ScrLB1/GUSR and LOXP/ScrRB1 for *OsDDM1b* represent primers for PCR analysis. Primers F8010/R9010, ScrLB1/GUSR, and SeqN2/SeqB2 were used for PCR genotyping of the WT or knock-out alleles. Horizontal boxes with 5', HPT, 3' and GUS indicate DNA probes for Southern blot hybridization. The size of restriction fragments for

Southern blot hybridization are also indicated. Restriction sites: B, *Bam*HI; V, *Eco*RV; H, *Hind*III; Nco, *Nco*I; S, *Sal*I; Sma, *Sma*I; X, *Xba*I.

```
OsDDM1a -----MSEVRSVPGMKVEASPAAAAAATVDSPTSVLEDDEISECKNGDVLDTTEAIKQEE
OsDDM1b MVEGMVVAVANGTKVEAAPAAGAGAAANADSPTSVLED-EISECKNGDASDTVEAIKQED
DDM1    --MVSLRSRKVIPASEMVSDGKTEKDASGDSPTSVLNEEENCEEKSVTVVEEEILLAK--
             :      .  :  .   :   *. *******:: * .* *. .  :    : :
OsDDM1a DHLDVLIEEKVDGFVDASSSLNVEPAANNSDLSPLTVPVKEEGQLLEPVKEEKADDFVDA
OsDDM1b DHLKVLAEERADDFVDASSSLTVELAANNGDLSPHPVPVKEEDQLLEPVKEEKADDFVDA
DDM1    ------------------------------------------------------------
OsDDM1a VPSLPIDLEAKNGDASLITDAMKEEEEKLHDARVKAEEEEVARKREEAARLAFDPNARFN
OsDDM1b TSSLPIDLEAKNGDASLITDAMKEEEDKLHEARVKADEEEEARKRAEAARLAFDPNARFN
DDM1    -----------NGDSSLISEAMAQEEE----QLLKLREDEEKANNAGSAVAPNLNETQFT
                  ***:***::** :**:      :*  *:*   :.  :*  .   :::*.
OsDDM1a KLDELLSQTQLYSEFLLEKMETIADVEGVQTHAEEEPVEEKKNGRGRKRKATSAPKYNDK
OsDDM1b KLDELLTQTQLYSEFLLEKMETIADVEGVDTPDEEEPVEEKKKGRGRKRKATSAPKYNNK
DDM1    KLDELLTQTQLYSEFLLEKMEDITING-----IESESQKAEPEKTGRGRKRKAASQYNNT
        ******:************** *:         *.*. : : :  ** ** .:*.:**:.
OsDDM1a KAKKAVAVMLTRSHEDCSPEDCTLTEEERWEKEQARLVPLMTGGKLKSYQIKGVKWLISL
OsDDM1b KAKKAVAAMLTRSREDCSPEDCTLTEEERWEKEQARLVPLMTGGKLKSYQIKGVKWLISL
DDM1    KAKRAVAAMISRSKEDGETINSDLTEEETVIKLQNELCPLLTGGQLKSYQLKGVKWLISL
        ***:***.*::**:** .. :. *****   * * .* **:***:*****:*********
OsDDM1a WQNGLNGILADQMGLGKTIQTIGFLAHLKGKGLDGPYLIIAPLSTLSNWVNEISRFVPSM
OsDDM1b WQNGLNGILADQMGLGKTIQTIGFLAHLKGKGLDGPYLIIAPLSTLSNWVNEISRFVPSM
DDM1    WQNGLNGILADQMGLGKTIQTIGFLSHLKGNGLDGPYLVIAPLSTLSNWFNEIARFTPSI
        **************************:****:********:**********.***:**.**:
OsDDM1a TGLIYHGDKAARAEIRRKFMPKTTGPDFPLIVTSYEMAMSDAK-HLAHYKWKYVIVDEGH
OsDDM1b TGVIYHGDKAARAEIRRKFMPKTTGPNFPLIVTSYEMAMSDAK-QLAHYKWKYVIVDEGH
DDM1    NAIIYHGDKNQRDELRRKHMPKTVGPKFPIVITSYEVAMNDAKRILRHYPWKYVVIDEGH
        ..:******  * *:***.****.**.**:::****:**.***  * ** ****::****
OsDDM1a RLKNSKCLLLRELKRLPMDNKLLLTGTPLQNNLAELWSLLNFILPDIFSSHQEFESWFDF
OsDDM1b RLKNSKCILLRELKRLPMDNKLLLTGTPLQNNLAELWSLLNFILPDIFSSHQEFESWFDF
DDM1    RLKNHKCKLLRELKHLKMDNKLLLTGTPLQNNLSELWSLLNFILPDIFTSHDEFESWFDF
        **** ** ******:* ***************:***************:**:********
OsDDM1a SAKGGEEEQEDSEEKRKVDVVSKLHAILRPFLLRRMKEDVEHMLPRKKEIIIYANMTDHQ
OsDDM1b CAKGGEEEQEESEEKRKVDVVSKLHAILRPFLLRRMKEDVEHMLPRKKEIIIYANMTNHQ
DDM1    SEKNKNEATKEEEEKRRAQVVSKLHGILRPFILRRMKCDVELSLPRKKEIIMYATMTDHQ
        . *. :*  ::.****:.:******.*****:***** ***  ********:**.**:**
OsDDM1a KQIQNHLVEQTFDQYLHEKSEIVLRKPGIKAKLNNLLIQLRKNCNHPDLLESAYDSSGLY
OsDDM1b KEIQNHLVEQTFDEYLHEKSEIVLRRPGIKAKLNNLLIQLRKNCNHPDLLESAYDSSGMY
DDM1    KKFQEHLVNNTLEAHLGEN---AIRGQGWKGKLNNLVIQLRKNCNHPDLLQGQIDGSYLY
        *::*:***::*:: :* *:   .:*  * *.*****:**************:.  *.* :*
OsDDM1a PPVEKLLEQCGKFQLLNRLLSLLLARKHKVLIFSQWTKVLDIIEYYLETKGLQVCRIDGS
OsDDM1b PPVEKLLEQCGKFQLLNRLLNLLLARKHKVLIFSQWTKVLDIIEYYLETKGLQVCRIDGS
DDM1    PPVEEIVGQCGKFRLLERLLVRLFANNHKVLIFSQWTKLLDIMDYYFSEKGFEVCRIDGS
        ****::: *****:**:***  *:*.:***********:***::**:. **::*******
OsDDM1a VKLEERRRQIAEFNDLNSSMNIFILSTRAGGLGINLTSADTCILYDSDWNPQMDLQAMDR
OsDDM1b VKLEERRRQIAEFNDLNSSMNIFILSTRAGGLGINLTSADTCILYDSDWNPQMDLQAMDR
DDM1    VKLDERRRQIKDFSDEKSSCSIFLLSTRAGGLGINLTAADTCILYDSDWNPQMDLQAMDR
        ***:****** :*.* :** .**:**************:**********************
OsDDM1a CHRIGQTRPVHVYRLATSHSVEGRIIKKAFGKLRLEHVVIGKGQFEQDRAKPN-ALDEAE
OsDDM1b CHRIGQTRPVHVYRLATSHSVEGRIIKKAFGKLRLEHVVIGKGQFEQDRAKPN-ALDEAE
DDM1    CHRIGQTKPVHVYRLSTAQSIETRVLKRAYSKLKLEHVVIGQGQFHQERAKSSTPLEEED
        *******:*******:*::*:* *::*:*:.**:********:***.*:***.. .*:* :
OsDDM1a LLALLRDEQGDEDRMIQTDISDEDLLKVMDRSDLTG-PPANADAAPLVPLKGPGWEVVVP
OsDDM1b LLALLRDEQDEEDRMIQTDITDEDLLKVMDRSDLTG-PPANADAAPLVPLKGPGWEVVVP
DDM1    ILALLKEDETAEDKLIQTDISDADLDRLLDRSDLTITAPGETQAAEAFPVKGPGWEVVLP
        :*********  **::******:* ** :::******  .*.:::**  .*:*********:*
OsDDM1a TKSGGGMLTSLTS
OsDDM1b TKSGGGMLTSLTS
DDM1    SS--GGMLSSLNS
        :.  ****:**.*
```

Figure 2

Figure 2. Amino acid sequence alignment of OsDDM1a and OsDDM1b with DDM1 of Arabidopsis.

Positions of amino acid identity, identical amino acids (*), conservative amino acid changes (:), and semi-conservative amino acid substitutions (.), among all three sequences are shown. Gaps are also indicated (-).

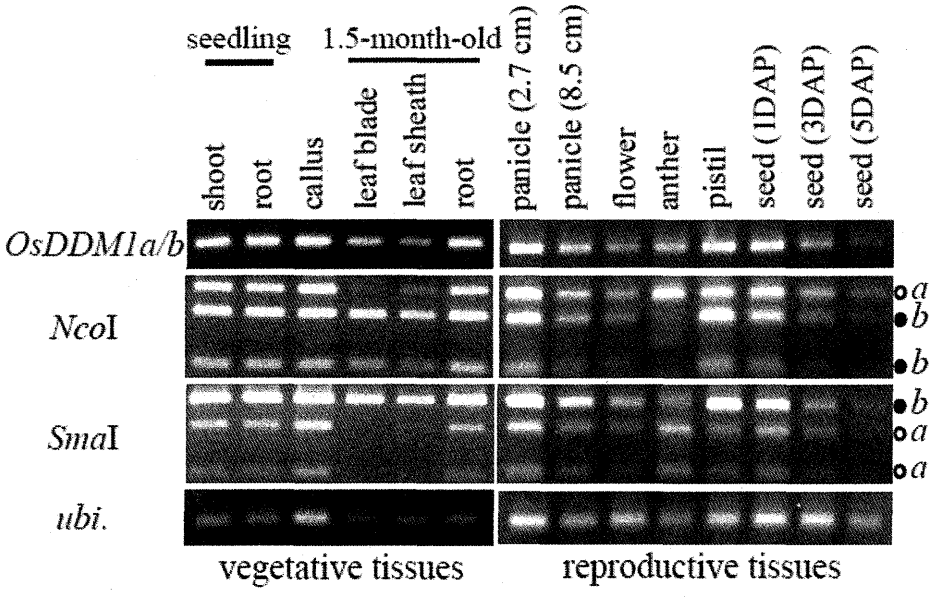

seedling
1.5-month-old
shoot
root
callus
leaf blade
leaf sheath
root
panicle (2.7 cm)
panicle (8.5 cm)
flower
anther
pistil
seed (1DAP)
seed (3DAP)
seed (5DAP)
OsDDM1a/b
NcoI
SmaI
ubi.
a
b
b
b
a
a
vegetative tissues
reproductive tissues

Figure 3

Figure 3. The expression of *OsDDM1a* and *OsDDM1b* detected by RT-PCR analysis in various tissues.

The expression of both *OsDDM1a* and *OsDDM1b* was detected by RT-PCR analysis in various tissues. RT-PCR products were digested by *Nco*I or *Sma*I. Transcripts of *OsDDM1a* and *OsDDM1b* were distinguished from each other by length after digestion. Open and filled circles correspond to *OsDDM1a* and *OsDDM1b*, respectively. Days after pollination (DAP).

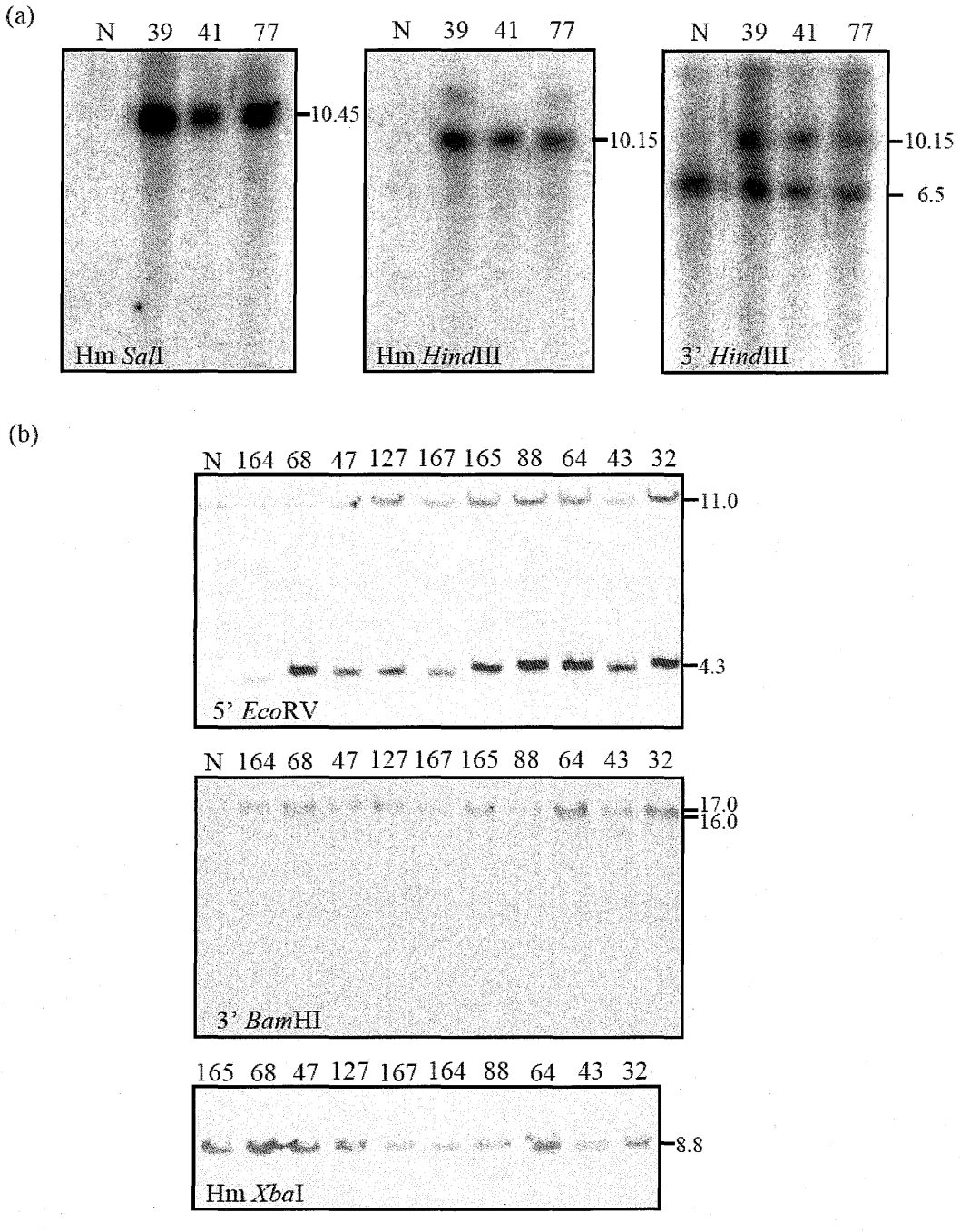

(a)
N 39 41 77
10.45
Hm SalI
N 39 41 77
10.15
Hm HindIII
N 39 41 77
10.15
6.5
3' HindIII
(b)
N 164 68 47 127 167 165 88 64 43 32
11.0
4.3
5' EcoRV
N 164 68 47 127 167 165 88 64 43 32
17.0
16.0
3' BamHI
165 68 47 127 167 164 88 64 43 32
8.8
Hm XbaI

Figure 4

Figure 4. Structural analysis of the targeted locus.

(a) and (b) Southern blot analysis of primary (T_0) targeted rice plants for *OsDDM1a* (a) and *OsDDM1b* (b). The restriction enzymes to cleave the genomic DNA samples and the probes used are indicated without and within brackets, respectively in each panel; N, Nipponbare.

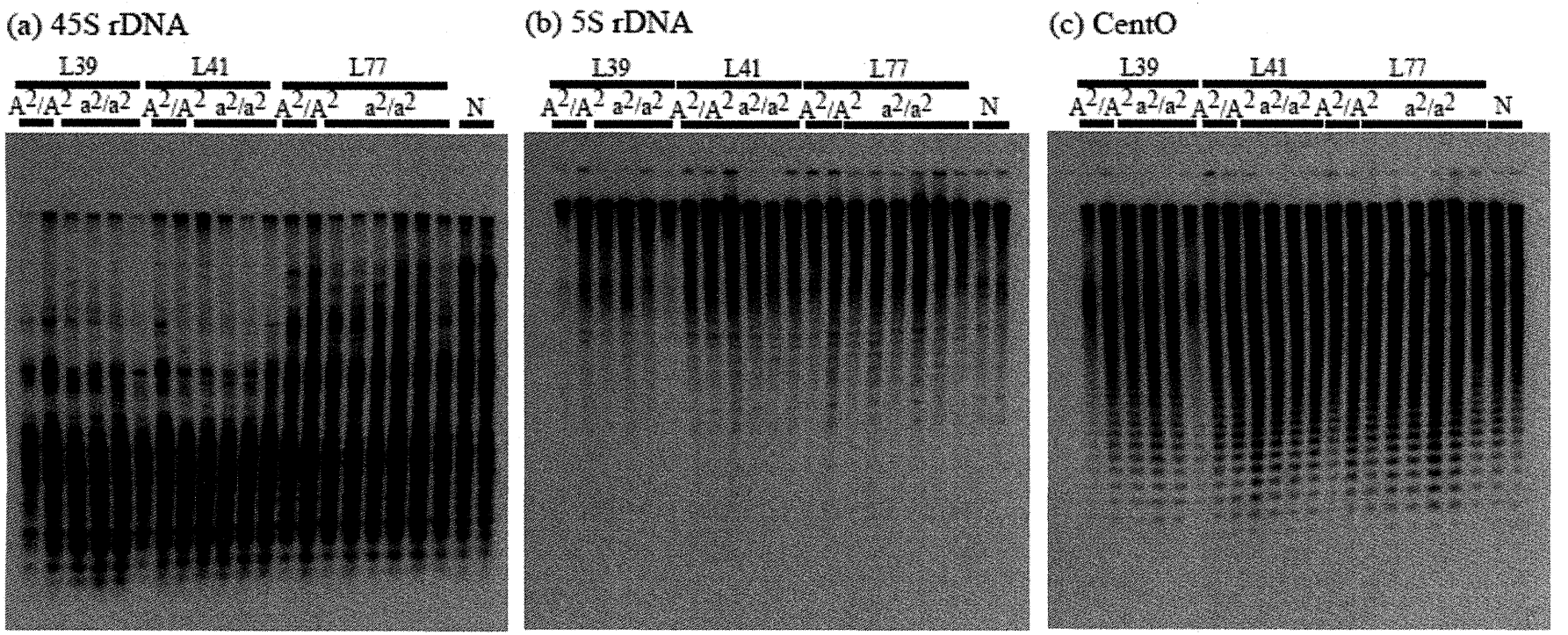

(a) 45S rDNA
L39
L41
L77
A2/A2 a2/a2 A2/A2 a2/a2 A2/A2 a2/a2 N
(b) 5S rDNA
L39
L41
L77
A2/A2 a2/a2 A2/A2 a2/a2 A2/A2 a2/a2 N
(c) CentO
L39
L41
L77
A2/A2 a2/a2 A2/A2 a2/a2 A2/A2 a2/a2 N

Figure 5

Figure 5. Effect of *OsDDM1a* disruption on DNA methylation levels in repetitive sequences.

Genomic DNA samples were prepared from the 16 or 17th leaves (right before flag leaves) of T_2 plants and those from the control Nipponbare plant (N), and digested with methylation-sensitive restriction enzyme, *Hpa*II. *Hpa*II recognizes CCGG motif but will not cut if the inner or outer cytosine is methylated. The *Hpa*II fragments of genomic DNA were hybridized to the probes; 45S rDNA (a), 5S rDNA (b), and CentO (c). The symbols a^2/a^2 and A^2/A^2 represent T_2 plants produced by selfing T_1 *Osddm1a* mutant plant (a/a) and T_1 WT plant (A/A), respectively, which are segregants in the selfed progeny of T_0 plants.

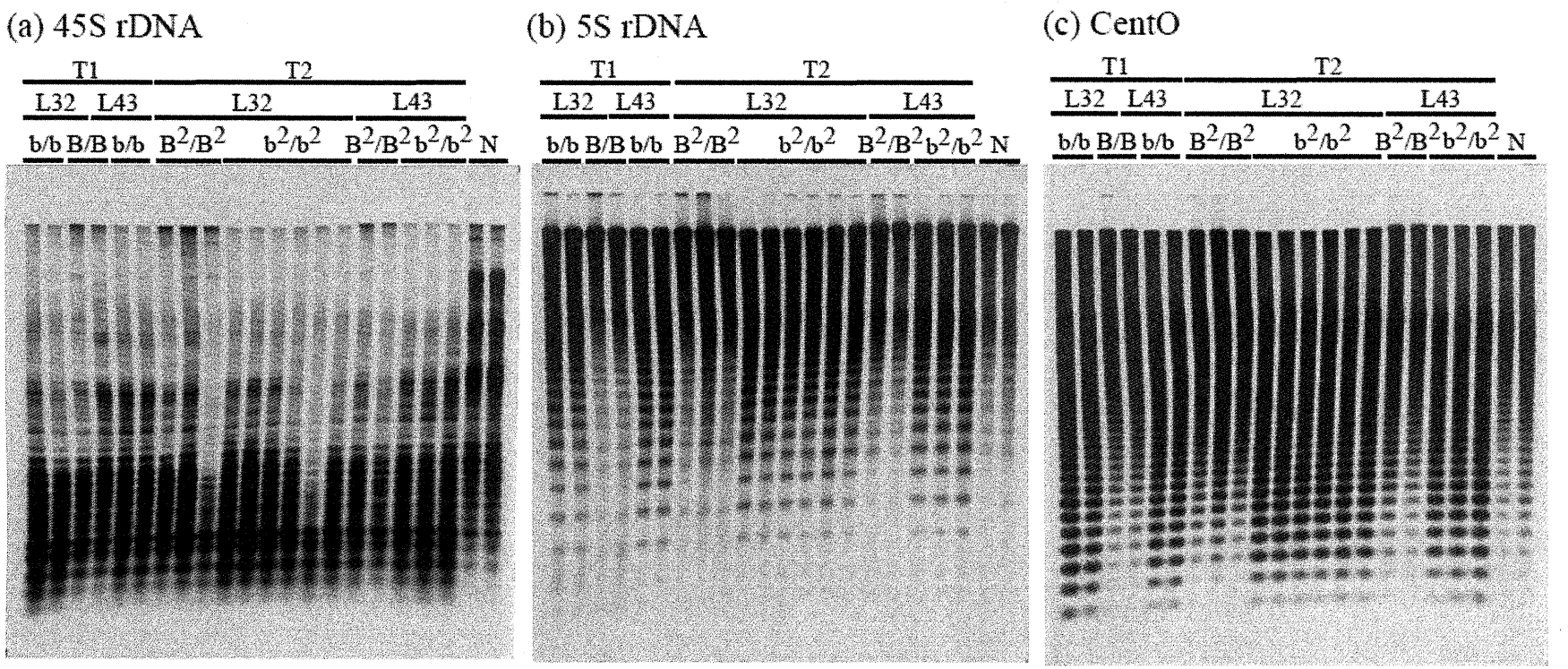

(a) 45S rDNA
T1
T2
L32
L43
L32
L43
b/b
B/B
b/b
B2/B2
b2/b2
B2/B2b2/b2
N
(b) 5S rDNA
T1
T2
L32
L43
L32
L43
b/b
B/B
b/b
B2/B2
b2/b2
B2/B2b2/b2
N
(c) CentO
T1
T2
L32
L43
L32
L43
b/b
B/B
b/b
B2/B2
b2/b2
B2/B2b2/b2
N

Figure 6

Figure 6. Effect of *OsDDM1b* disruption on DNA methylation levels in repetitive sequences.

Genomic DNA samples were prepared, digested by *Hpa*II, and hybridized as described in Figure 5. The symbols b/b and B/B are used as described in Figure 5 in the selfed progeny of the heterogously targeted T_0 plants, whereas b^2/b^2 and B^2/B^2 represent T_2 plants obtained by selfing a T_1 mutant plant (b/b) and a T_1 WT plant (B/B), respectively.

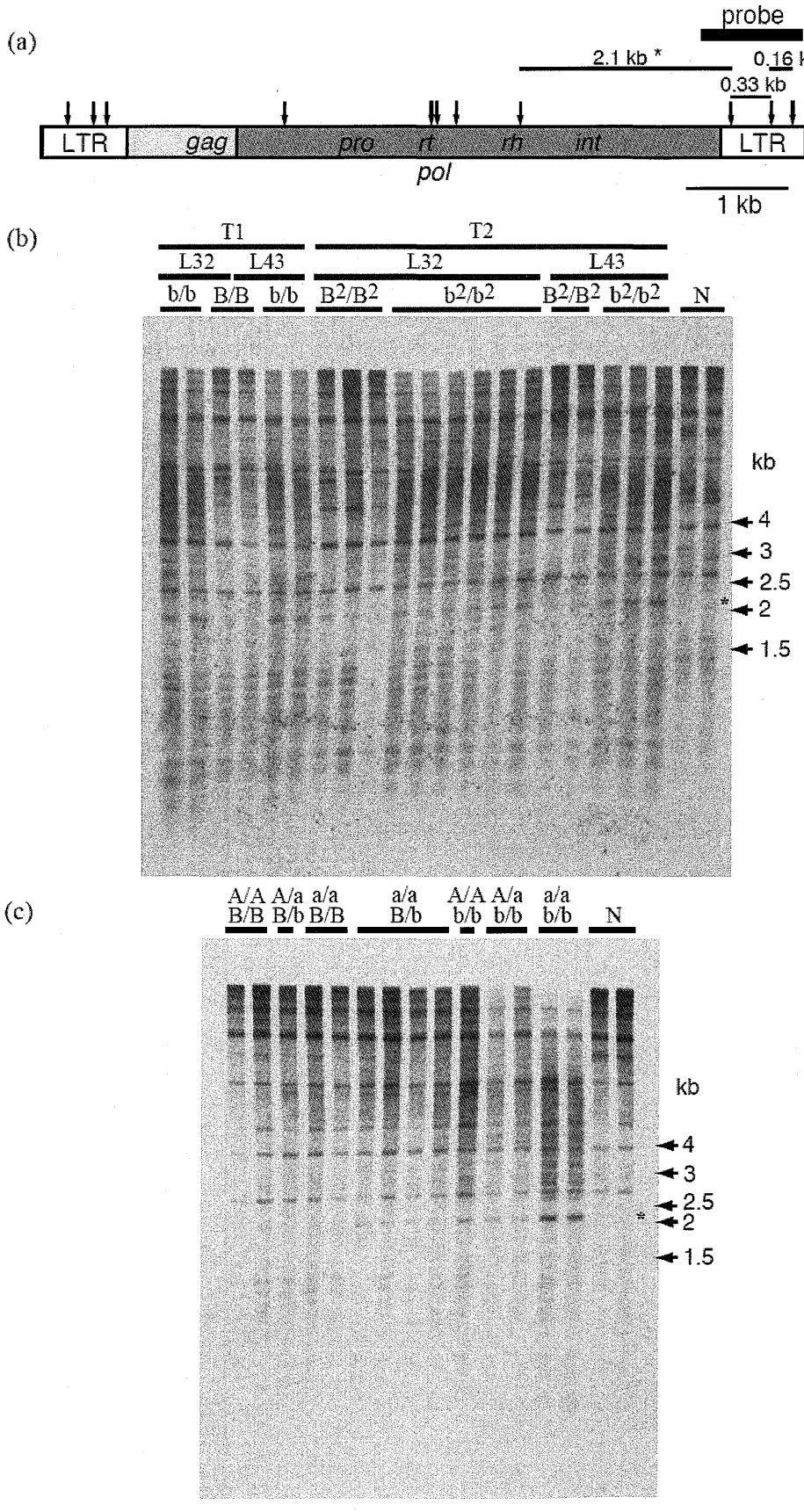

(a)
probe
2.1 kb *
0.16 kb
0.33 kb
LTR
gag
pro
rt
rh
int
LTR
pol
1 kb
(b)
T1
T2
L32
L43
L32
L43
b/b
B/B
b/b
B2/B2
b2/b2
B2/B2
b2/b2
N
kb
4
3
2.5
2
1.5
(c)
A/A B/B
A/a B/b
a/a B/B
a/a B/b
A/A b/b
A/a b/b
a/a b/b
N
kb
4
3
2.5
2
1.5

Figure 7

Figure 7. Analysis of DNA methylation levels in *RIRE7*.

(a) Schematic structure of *RIRE7*. LTR is shown by open box, and *gag* and *pol* regions are shown by light and dark gray boxes, respectively. The *gag* gene encodes the viral envelope Gag protein, and *pol* comprises *pro*, *rt*, *rh*, and *int* genes, which encode proteinase, reverse transcriptase, RNaseH, and integrase, respectively. The horizontal dark-gray bar with probe indicates DNA probe for Southern blot analysis and the vertical arrows indicate *Hpa*II sites. The horizontal black bar with asterisk indicates the *Hpa*II fragment that is corresponding to the bands indicated with asterisk on each panel in (b) and (c).

(b) and (c) Southern blot analysis in T_1 and T_2 *Osddm1b* mutant plants (b) and in progenies of the selfed *OsDDM1a* and *OsDDM1b* double hetero (*OsDDM1a/Osddm1a*, *OsDM1b/Osddm1b*) plant (c). Genomic DNA samples were prepared, digested by *Hpa*II, and hybridized to *RIRE7* LTR probe shown in (a). Symbols are used as described in Figures 5 and 6.

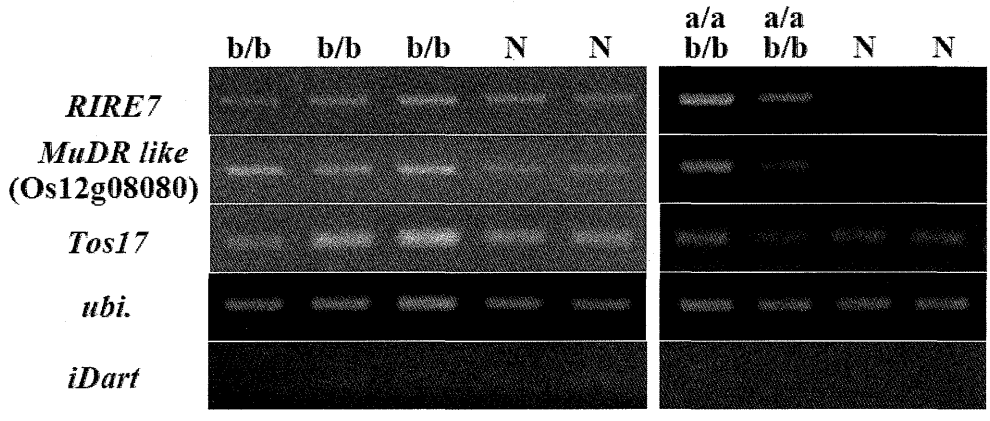

b/b
b/b
b/b
N
N
a/a
b/b
a/a
b/b
N
N
RIRE7
MuDR like
(Os12g08080)
Tos17
ubi.
iDart

Figure 8

Figure 8. Transcriptional reactivation of TEs in *Osddm1* mutant plants.

RT-PCR analysis of the selected TEs was performed. Symbols are used as described in Figures 5 and 6.

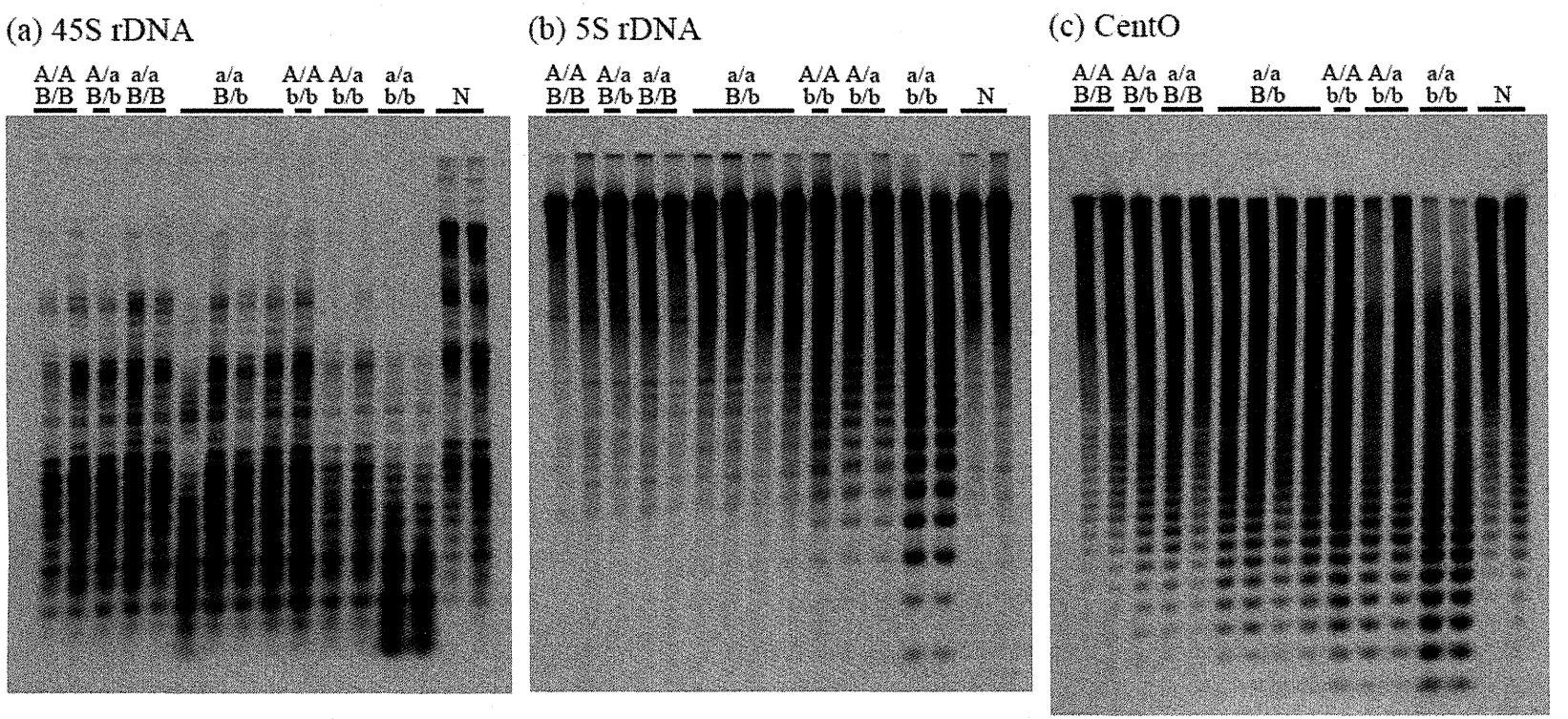

(a) 45S rDNA
A/A B/B
A/a B/b
a/a B/B
a/a B/b
A/A b/b
A/a b/b
a/a b/b
N
(b) 5S rDNA
A/A B/B
A/a B/b
a/a B/B
a/a B/b
A/A b/b
A/a b/b
a/a b/b
N
(c) CentO
A/A B/B
A/a B/b
a/a B/B
a/a B/b
A/A b/b
A/a b/b
a/a b/b
N

Figure 9

Figure 9. Analysis of DNA methylation levels in progenies of the selfed *OsDDM1a* and *OsDDM1b* double hetero (a/A, b/B) plant.

Genomic DNA samples were prepared, digested by *Hpa*II, and hybridized as described in Figure 5. The symbols are used as described in Figures 5 and 6.

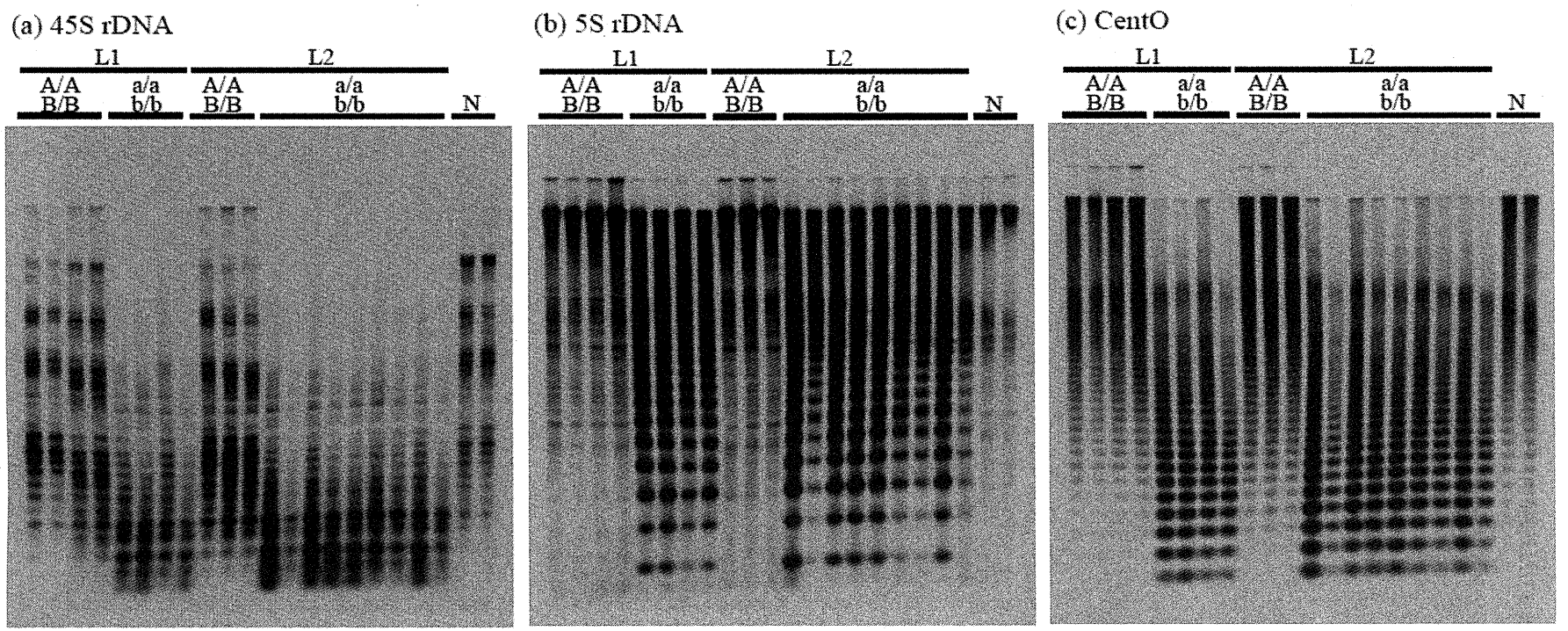

(a) 45S rDNA
L1
L2
A/A B/B
a/a b/b
A/A B/B
a/a b/b
N
(b) 5S rDNA
L1
L2
A/A B/B
a/a b/b
A/A B/B
a/a b/b
N
(c) CentO
L1
L2
A/A B/B
a/a b/b
A/A B/B
a/a b/b
N

Figure 10

Figure 10. Analysis of DNA methylation levels in several *Osddm1a Osddm1b* double mutant plants.

Genomic DNA samples were prepared, digested by *Hpa*II and hybridized as described in Figure 5. The symbols are used as described in Figures 5 and 6.

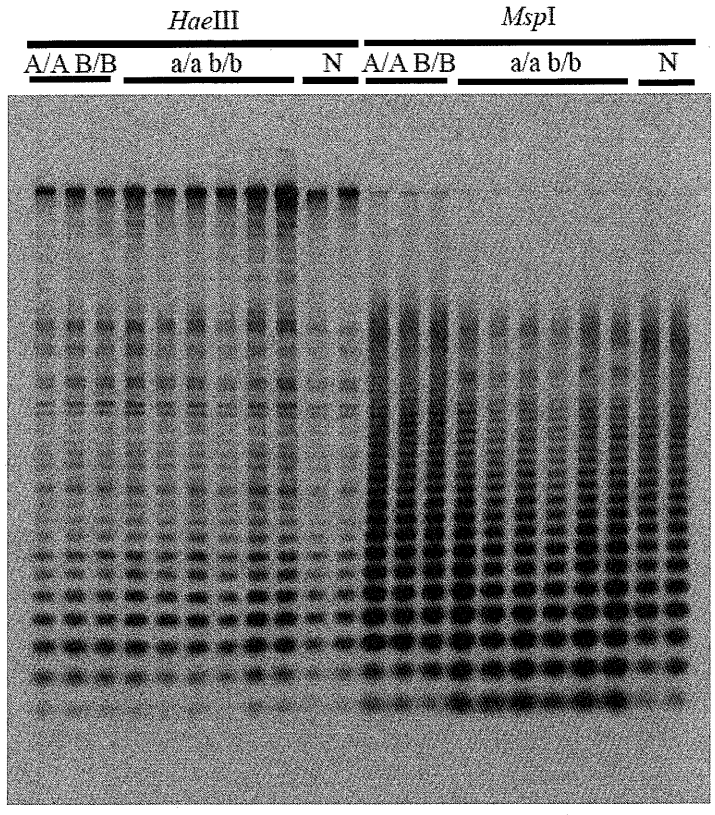


Figure 11

Figure 11. Analysis of non-CG (CHG and CHH) DNA methylation levels in *Osddm1a Osddm1b* double mutant plants.

Genomic DNA samples were prepared as described in Figure 5 and digested with methylation sensitive restriction enzymes, *Msp*I or *Hae*III. *Msp*I is an isoschizomer of *Hpa*II and cuts CCGG motifs but will not cut if the outer cytosine is methylated. *Hae*III recognizes GGCC motifs but will not cut if the inner cytosine is methylated. *Msp*I or *Hae*III fragments of genomic DNA were hybridized to the CentO probe. Symbols are used as described in Figures 5 and 6.

Table 1. List of primers.

Designation	Sequences	Designation	Sequences	Comments
Primer sets for PCR and RT-PCR amplification				
HindIII+DDM1a-1	AAGCTTGGAACAGAATAGTTCTGAAAAATTGTGCTA	SrfI+DDM1a-1 Rev	GCCCGGGCTCGAGAAAAAAAGTGGGCGAGAGAAAAAGA	5' homologous region of *OsDDM1a* in pJHYDDM1a (Figure 1c)
PmeI+DDM1a-2	GTTTAAACTCTTTTCTTTTCTTGGAGTGTCGAGTTCAC	AscI+DDM1a-2 Rev	GGCGCGCCGTTGATGGGAATATAATTGAAATTGCTGTT	3' homologous region of *OsDDM1a* in pJHYDDM1a (Figure 1c)
AscI+DDM1a-2 contF	GGCGCGCCTAACCCTGTCTATATATATGTTATTCC	I-SceI+DDM1a-2 contR	ATTACCCTGTTATCCCTAGTAATCCCTGAATTAAACAGAATAGG	Additional 3' homologous region of OsDDM1a in pJHYDDM1aC
5F PacI	GTTAATTAAGTCTGATCCTATTCTGTTTGG	5R HindIII	TCTAAGCTTCTTCCTCTCCCCCTTCCGCT	5' homologous region of *OsDDM1b* in pJHY-Ki-DDM1b (Figure 1f)
C5F I-CeuI	CCCGTAACTATAACGGTCCTAAGGTAGCGAATTTCCGGATCGTGCCGTGCC	C5R PacI	GGTTAATTAACGGAGCTTCGTTTTTGAAAAGG	Additional 5' homologous region of *OsDDM1b* in pJHY-Ki-DDM1bC
3F PmeI	AGGTTTAAACAAGAGGGGTCGGATAGTAAG	3R AscI	TGGGCGCGCCGTGCTGACTGCTGAGACGTG	3' homologous region of *OsDDM1b* in pJHY-Ki-DDM1b (Figure 1f)
C3F AscI	TTGGCGCGCCTAAAGTACAAATATGATAGT	C3R I-SceI	CCCTATATTACCCTGTTATCCCTAGCGTAACTCATAGCATCCCAGTAAACACC	Additional 3' homologous region of *OsDDM1b* in pJHY-Ki-DDM1bC
DDM1a-F5160	CGCGACCACCCTCTTTATGAGCTCGTGACT	LoxR	GCATACATTATACGAAGTTATGCCCGGGC	5' junction fragment for *OsDDM1a* (Figure 1c)
LoxP30F	GTATAATGTATGCTATACGAAGTTATGTTT	I-SceI+DDM1a-2 contR	ATTACCCTGTTATCCCTAGTAATCCCTGAATTAAACAGAATAGG	3' junction fragment for *OsDDM1a* (Figure 1c)
ScrLB1	TGTGAATGGTGTGCATGTGG	GUSR	ATTCCACAGTTTTCGCGATC	5' junction fragment for *OsDDM1b* (Figure 1f)
LoxP	ATAACTTCGTATAGCATACATTATACGAAGTTAT	ScrRB1	ACACGTTAGCAGACATCGCA	3' junction fragment for *OsDDM1b* (Figure 1f)
F8010	CTCGTGGTTTCCAGACGAGTTATGAATTAG	R9010	GGTATCCAAGACATCCCCATTCTTACATTC	Genotyping for WT and knock-in allele for *OsDDM1a*
ScrLB1	TGTGAATGGTGTGCATGTGG	GUSR	ATTCCACAGTTTTCGCGATC	Genotyping for knock-in allele for *OsDDM1b*
SeqN2	TGGCATTCCTCGTAGATAAG	SeqB2	ACAAATCAGCCTCAGCGATC	Genotyping for WT for *OsDDM1b*
RT DDM F	GGGCTTAATGGGATATTGGCTGATC	RT DDM R	TATGCTCTACATCCTCCTTCATCCG	RT-PCR for *OsDDM1a* and *OsDDM1b* (Figure 1a,d; Figure 3)
HmF	GTAGGTAGACCGGGGGCAATGAG	HmR	GTCGCATCGACCCTGCGCCCAAGCTGCATCATCG	HPT probe (Figure 4a)
hygpF	CTCGTGCTTTCAGCTTCGATGTAGG	hygpR	GAAGATGTTGGCGACCTCGTATTGG	HPT probe (Figure 4b)
GUSdigF	ATTGCTGTGCCAGGCAG	GUSdigR	TAACCTTCACCCGGTTG	GUS probe (Figure 4b)
DDM1a-F4190	GAGTTCTAGCTTCATACATG	DDM1a-seq8	CACATTCATACAGAACACTG	3' region probe for *OsDDM1a* (Figure 4a)
DDM1b 5' labelF2	TTCAGGAACTGAACACACC	DDM1b 5' labelR3	ACCTCCTCGGAAGTCAGTG	5' region probe for *OsDDM1b* (Figure 4b)
DDM1b 3' labelF	CTGCTTCTTCTTCTTCTCTG	DDM1b 3' labelR	GCCAGTTCATACAATAGTGC	3' region probe for *OsDDM1b* (Figure 4b)
45S F	CAAGGTCCTGAGCACGTACCGCGCG	45S R	GGACCGAACCGGGTGCCGTTTGCGC	45S probe (Figure 5a; Figure 6a; Figure 9a; Figure 10a)
5S F	GATCCCATCAGAACTCCGAAG	5S R	GGTGCTTTAGTGCTGGTATG	5S probe (Figure 5b; Figure 6b; Figure 9b; Figure 10b)
centO F	CGTTCGTGGCAAAAACTCAC	centO R	CGAACGCACCCAATACACTC	CentO probe (Figure 5c; Figure 6c; Figure 9c; Figure 10c; Figure 11)
RIRE7 F	GGACAAGAGATATTCAGTGCCTACG	RIRE7 R	CAGCTATGTATGGCTTCAACCTTGG	RIRE7 probe (Figure 7b,c)
RIRE7 RT F	CACAAATTGCAACAACTTAGAC	RIRE7 RT R	AATCTCCTTTGTTTCCTCCG	RT-PCR for TE derepression (Figure 8)
MuDR RT F	GAGGATATGTCATTGAAGACC	MuDR RT R	CACATCATCACTACCGCAATG	RT-PCR for TE derepression (Figure 8)
Tos17 RT F	CACCAGGTGTGGAAGCTCCAC	Tos 17 RT R	TACCACTGAGCTGAAGCGTGC	RT-PCR for TE derepression (Figure 8)
Dart RT F	CGTAGTTCAACAGTTTGATCGCAGAGG	Dart RT R	CTACTGGGGTTGGTCTGG	RT-PCR for TE derepression (Figure 8)
Ubi RT F	CCAGGACAAGATGATCTGCC	Ubi RT R	AAGAAGCTGAAGCATCCAGC	RT-PCR for TE derepression (Figure 8)

Table 2. Up-regulated genes in *Osddm1s* mutant seedlings.

ProbeName*	mut1/N1 Ratio	mut2/N1 Ratio	mut1-2/N2 Ratio	mut2-2/N2 Ratio	Description**
Os01g0108500	13.13	7.59	22.60	18.17	(No Hit)
Os01g0113900	9.05	30.47	7.32	16.28	Conserved hypothetical protein.
Os01g0117200	3.01	8.65	9.08	5.23	ARK protein (Fragment).
Os01g0118400	2.84	6.21	2.69	2.65	HGA6.
Os01g0121000	18.05	3.28	4.88	6.48	(No Hit)
Os01g0124600	15.70	17.14	46.67	17.42	(No Hit)
Os01g0159200	10.81	12.04	9.20	2.76	Arabinogalactan protein.
Os01g0166500	9.21	5.73	4.28	25.92	Conserved hypothetical protein.
Os01g0175900	15.39	14.77	7.67	9.84	Conserved hypothetical protein.
Os01g0185300	2.13	2.72	2.02	2.10	Transferase family protein.
Os01g0190900***	2.10	6.17	2.09	3.46	Mutator-like transposase.
Os01g0214600	15.11	6.13	30.62	72.84	Lipolytic enzyme, G-D-S-L family protein.
Os01g0288700	4426.89	3065.99	5089.34	2834.43	Non-protein coding transcript, uncharacterized transcript.
Os01g0293000	4.09	6.00	6.17	2.21	S-adenosylmethionine synthetase 1 (EC 2.5.1.6) (Methionine adenosyltransferase 1) (AdoMet synthetase 1).
Os01g0307300	3.01	2.44	7.04	12.15	(No Hit)
Os01g0329000	8.94	6.74	2.04	3.12	Phospholipid/glycerol acyltransferase family protein.
Os01g0346900	14.07	29.86	7.60	22.79	Histone-fold domain containing protein.
Os01g0356800	4.05	10.18	12.82	59.86	Peptidase S9A, prolyl oligopeptidase, N-terminal beta-propeller domain containing protein.
Os01g0356900	7.70	2.87	12.46	17.56	Extensin-like protein.
Os01g0503700	2.52	6.54	3.56	4.45	(No Hit)
Os01g0549300	16.55	1837.28	79.31	1938.12	Homeodomain-like containing protein.
Os01g0554900	3.01	2.39	7.00	11.81	(No Hit)
Os01g0566500	39.36	36.27	2.55	3.56	Dioxygenase RAMOSUS1.
Os01g0567200	23.37	17.58	12.34	11.70	Conserved hypothetical protein.
Os01g0586700	10.24	3.50	26.03	2.20	(No Hit)
Os01g0611000	2.02	4.44	2.86	67.88	Protein of unknown function

					DUF642 family protein.
Os01g0666000	4.40	8.71	4.10	3.23	Lipid phosphate phosphatase 2 (EC 3.1.3.-) (AtLPP2) (Phosphatidic acid phosphatase 2) (AtPAP2) (Prenyl diphosphate phosphatase).
Os01g0678000	3.99	7.85	5.99	5.57	Conserved hypothetical protein.
Os01g0692400	144.15	166.74	10.28	14.54	Conserved hypothetical protein.
Os01g0700900	102.06	157.66	110.11	44.99	Cytochrome P450 family protein.
Os01g0701400	18.38	34.06	64.26	23.64	Cytochrome P450 family protein.
Os01g0705200	8.99	9.73	2.41	8.72	Late embryogenesis abundant protein repeat containing protein.
Os01g0705200	7.23	10.65	2.61	8.95	Late embryogenesis abundant protein repeat containing protein.
Os01g0705200	10.61	10.20	2.34	8.07	Late embryogenesis abundant protein repeat containing protein.
Os01g0717000	2.99	3.22	2.94	2.04	GmCK1p (EC 2.7.1.32).
Os01g0717000	3.05	3.21	3.26	2.20	GmCK1p (EC 2.7.1.32).
Os01g0717000	2.83	3.06	2.82	2.06	GmCK1p (EC 2.7.1.32).
Os01g0717000	3.04	3.00	3.27	2.13	GmCK1p (EC 2.7.1.32).
Os01g0778600	2.34	16.51	4.33	2.74	(No Hit)
Os01g0779500	3.68	9.75	3.04	3.81	(No Hit)
Os01g0810400	2.71	6.40	3.65	4.61	(No Hit)
Os01g0829300	392.12	19.68	22.74	2.20	(No Hit)
Os01g0858200	5.14	2.86	6.91	2.21	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 16.
Os01g0862800	15.87	9.39	11.92	2.74	No apical meristem (NAM) protein domain containing protein.
Os01g0866200	2.88	2.11	2.28	2.41	Histone H3.
Os01g0869400	4.09	2.81	2.02	2.67	UDP-glucuronosyl/UDP-glucosyltransferase family protein.
Os01g0907800	3.54	9.11	3.95	4.97	(No Hit)
Os01g0931000	5.11	10.26	3.90	3.24	Glucose/ribitol dehydrogenase family protein.
Os02g0118300	3.06	7.93	3.18	3.80	(No Hit)
Os02g0140500	2.14	5.76	2.04	2.70	mutator-like transposase [Oryza sativa (japonica cultivar-group)].
Os02g0161100	4.93	9.12	21.62	53.43	Actin-binding FH2 domain containing protein.
Os02g0216200	7.69	12.56	17.28	19.18	Hypothetical protein.

Os02g0237100	2.24	19.27	8.18	31.85	Spermine synthase family protein.
Os02g0239400	5.54	14.16	3.48	15.15	(No Hit)
Os02g0240100	2.50	3.11	2.48	3.92	Peroxidase (EC 1.11.1.7).
Os02g0253700	2.13	5.87	12.56	12.32	Conserved hypothetical protein.
Os02g0255400	30.35	38.97	12.49	38.05	RNA-directed DNA polymerase (Reverse transcriptase) domain containing protein.
Os02g0280400	2.81	2.53	6.29	5.69	Protein kinase-like protein.
Os02g0300000	5.30	53.31	16.04	109.79	Conserved hypothetical protein.
Os02g0441000	6.88	11.41	3.57	5.08	Hypothetical protein.
Os02g0448000	25.24	21.87	20.94	29.81	Protein of unknown function DUF296 domain containing protein.
Os02g0481400	13.22	27.43	12.91	32.61	(No Hit)
Os02g0517900	3.37	5.89	31.40	7.84	Conserved hypothetical protein.
Os02g0547100	45.93	99.97	24.53	50.71	SWIM Zn-finger domain containing protein.
Os02g0577500	12.93	26.62	13.67	33.84	(No Hit)
Os02g0615500	12.80	39.43	5.35	2.27	Protein kinase domain containing protein.
Os02g0659900	4.74	12.65	4.97	14.72	(No Hit)
Os02g0665300	3.92	2.63	4.15	6.10	Proteinase inhibitor I9, subtilisin propeptide domain containing protein.
Os02g0671500	2.91	2.35	7.19	11.85	(No Hit)
Os02g0677300	3.89	2.14	15.00	2.17	CRT/DRE binding factor 1.
Os02g0677300	4.84	2.82	19.85	3.40	CRT/DRE binding factor 1.
Os02g0730000	7.47	10.73	2.54	3.84	Mitochondrial aldehyde dehydrogenase ALDH2a.
Os02g0772100	333.94	183.69	345.41	292.27	Conserved hypothetical protein.
Os02g0789700	75.79	63.79	71.66	73.10	Conserved hypothetical protein.
Os02g0822900	5.91	5.01	3.32	3.63	Protein kinase domain containing protein.
Os03g0115400	2.12	2.56	2.91	3.97	WD40-like domain containing protein.
Os03g0115700	24.56	30.30	37.48	33.22	Short-chain dehydrogenase/reductase SDR family protein.
Os03g0115800	4972.22	4797.53	5578.59	5410.22	Conserved hypothetical protein.
Os03g0115800	5787.49	5491.30	7198.69	7033.70	Conserved hypothetical protein.

Os03g0141100	4.58	8.33	6.29	17.78	Conserved hypothetical protein.
Os03g0168100	6.01	5.54	2.54	3.74	Late embryogenesis abundant protein repeat containing protein.
Os03g0183500	3.27	3.89	13.32	3.61	Protein of unknown function DUF581 family protein.
Os03g0233000	10.73	21.85	6.00	12.10	Protein of unknown function DUF607 family protein.
Os03g0233700	429.63	1166.49	664.17	1790.56	(No Hit)
Os03g0246500	3.12	2.16	2.82	7.10	Plant protein of unknown function DUF869 family protein.
Os03g0246500	3.43	2.22	3.15	7.56	Plant protein of unknown function DUF869 family protein.
Os03g0275700	3.35	2.42	3.18	2.58	Hypothetical protein.
Os03g0282300	10.28	9.67	21.49	17.21	Conserved hypothetical protein.
Os03g0303100	11.75	6.19	7.99	9.78	Hypothetical protein.
Os03g0304800	6.30	3.34	2.98	3.27	Conserved hypothetical protein.
Os03g0307200	6.87	6.29	6.10	5.58	Nicotianamine synthase 1 (EC 2.5.1.43) (S-adenosyl-L-methionine:S-adenosyl-L-methionine:S-adenosyl -methionine 3-amino-3-carboxypropyltransferase 1) (HvNAS1).
Os03g0327600	2.44	4.39	2.23	3.47	Ricin B-related lectin domain containing protein.
Os03g0327700	8.00	18.85	8.18	21.72	Hypothetical protein.
Os03g0345500	3.36	8.79	3.81	4.62	(No Hit)
Os03g0365000	4.41	10.62	6.02	10.96	(No Hit)
Os03g0365200	5.16	11.41	6.33	11.77	Peptidase C48, SUMO/Sentrin/Ubl1 family protein.
Os03g0429900	15.99	48.17	28.31	34.73	Nucleic acid-binding OB-fold domain containing protein.
Os03g0446000	27.90	18.56	18.60	16.92	(No Hit)
Os03g0575200	9.53	3.61	8.74	2.60	K+ potassium transporter family protein.
Os03g0575200	9.70	4.38	9.39	3.48	K+ potassium transporter family protein.
Os03g0589500	80.35	207.32	367.87	348.96	Conserved hypothetical protein.
Os03g0598500	4.49	10.12	5.65	10.52	(No Hit)
Os03g0629800	391.06	244.24	128.87	109.18	Conserved hypothetical protein.

Os03g0703100	6.88	7.91	2.31	16.05	Beta-glucosidase.
Os03g0714800	342.17	362.29	3.52	3.91	En/Spm-like transposon proteins family protein.
Os03g0717000	5.28	2.56	3.66	2.06	TMK protein precursor.
Os03g0765400	3.13	3.35	3.24	5.76	Conserved hypothetical protein.
Os03g0790100	34.12	71.55	3.36	24.64	(No Hit)
Os03g0800400	2.25	3.76	2.12	2.24	Conserved hypothetical protein.
Os03g0835200	4.47	15.38	5.78	9.18	Hypothetical protein.
Os03g0835200	2.76	6.85	3.69	5.73	Hypothetical protein.
Os03g0839800	4.42	2.27	3.86	5.06	Conserved hypothetical protein.
Os03g0841400	3.24	2.83	8.81	15.75	(No Hit)
Os03g0854900	14.22	25.88	13.66	32.08	(No Hit)
Os03g0858600	5.15	3.52	4.30	2.19	Protein of unknown function DUF668 family protein.
Os04g0108900	2.95	2.46	7.05	12.52	(No Hit)
Os04g0108900	3.05	2.70	7.58	13.03	(No Hit)
Os04g0115200	13.02	45.53	211.54	29.68	Conserved hypothetical protein.
Os04g0136600	2.96	5.28	2.07	2.63	(No Hit)
Os04g0141500	2.94	2.46	6.71	11.52	(No Hit)
Os04g0175600	4.25	3.97	2.41	2.25	Caffeic acid O-methyltransferase (EC 2.1.1.6).
Os04g0175600	4.31	4.39	2.44	2.21	Caffeic acid O-methyltransferase (EC 2.1.1.6).
Os04g0175900	2.76	2.17	4.46	4.92	0-methyltransferase (EC 2.1.1.6) (Fragment).
Os04g0175900	2.67	2.14	4.09	4.57	0-methyltransferase (EC 2.1.1.6) (Fragment).
Os04g0176200	2.42	4.03	4.54	4.13	S-adenosyl-L-methionine: beta-alanine N-methyltransferase (Fragment).
Os04g0192200	2.23	4.96	2.28	4.12	Conserved hypothetical protein.
Os04g0195800	2.95	2.50	6.83	11.90	(No Hit)
Os04g0206500	3.27	225.94	18.01	182.32	UDP-glucuronosyl/UDP-glucosyltransferase family protein.
Os04g0213400	2.97	2.48	6.94	11.93	(No Hit)
Os04g0219900	3.04	2.61	6.77	12.27	(No Hit)
Os04g0223600	9.79	19.76	2.13	3.22	Non-protein coding transcript, unclassifiable transcript.
Os04g0247700	13.22	20.71	14.11	33.00	Ribosomal protein L9 N-terminal-like domain containing

					protein.
Os04g0249600	2.57	2.11	2.13	4.22	Rhodanese-like domain containing protein.
Os04g0259800	6.41	4.93	3.63	7.71	Conserved hypothetical protein.
Os04g0268000	2.96	2.41	7.02	12.07	(No Hit)
Os04g0280700	12.88	25.89	11.66	29.06	(No Hit)
Os04g0295500	2.54	3.83	4.71	14.03	Conserved hypothetical protein.
Os04g0301500	6.58	8.99	25.39	5.02	Basic helix-loop-helix dimerisation region bHLH domain containing protein.
Os04g0326000	140.90	139.27	114.72	122.29	ARM repeat fold domain containing protein.
Os04g0326200	17.47	42.37	15.91	60.62	Ankyrin repeat containing protein.
Os04g0333700	2.90	2.42	7.39	12.98	(No Hit)
Os04g0340100	4.50	11.55	4.31	2.28	Protein kinase-like domain containing protein.
Os04g0344100	26.66	87.10	7.77	2.04	Terpene synthase-like domain containing protein.
Os04g0355900	3.56	12.86	3.55	20.68	(No Hit)
Os04g0368000	2.59	2.83	2.53	2.62	(No Hit)
Os04g0377400	4.61	11.87	6.09	11.82	(No Hit)
Os04g0397800	53.31	18.24	74.74	39.38	Non-protein coding transcript, unclassifiable transcript.
Os04g0397900	11.43	15.99	18.27	20.00	(No Hit)
Os04g0398000	7.81	9.09	3.59	2.04	Pathogenesis-related transcriptional factor and ERF domain containing protein.
Os04g0451600	3.06	2.53	6.73	11.80	(No Hit)
Os04g0459000	2.31	4.56	9.72	3.76	MDR-like ABC transporter.
Os04g0480200	2.42	6.95	5.02	2.19	Conserved hypothetical protein.
Os04g0497800	2.70	7.34	2.56	3.23	(No Hit)
Os04g0518900	3.68	8.92	3.49	4.07	(No Hit)
Os04g0575900	27.82	18.36	18.85	17.25	(No Hit)
Os04g0598600	2.36	3.18	2.06	2.12	Protein kinase domain containing protein.
Os04g0604000	777.31	162.96	1033.33	686.83	Actin filament bundling protein P-115-ABP.
Os04g0604000	822.39	187.81	1218.99	836.56	Actin filament bundling protein P-115-ABP.
Os04g0604300	8.41	2.56	9.05	8.79	Xyloglucan

					endotransglucosylase/hydrolase protein 24 precursor (EC 2.4.1.207) (At-XTH24) (XTH-24) (Meristem protein 5) (MERI-5 protein) (MERI5 protein) (Endo-xyloglucan transferase) (Xyloglucan endo-1,4-beta-D-glucanase).
Os04g0604300	8.66	2.73	10.10	10.04	Xyloglucan endotransglucosylase/hydrolase protein 24 precursor (EC 2.4.1.207) (At-XTH24) (XTH-24) (Meristem protein 5) (MERI-5 protein) (MERI5 protein) (Endo-xyloglucan transferase) (Xyloglucan endo-1,4-beta-D-glucanase).
Os04g0607800	7.60	4.19	30.68	24.35	(No Hit)
Os04g0677000	23.83	26.03	38.92	38.47	Conserved hypothetical protein.
Os04g0685200	2.76	3.14	2.30	3.19	Peptidase aspartic family protein.
Os04g0688200	84.31	152.47	27.87	31.63	Anionic peroxidase precursor.
Os05g0100100	35.28	64.26	14.96	14.44	(No Hit)
Os05g0100100	37.00	63.09	17.67	16.30	(No Hit)
Os05g0124000	9.05	15.67	9.74	58.78	Ankyrin repeat containing protein.
Os05g0124700	19.98	20.93	17.08	16.82	Conserved hypothetical protein.
Os05g0128100	2.74	3.89	2.15	3.28	Hypothetical protein.
Os05g0132700	8.08	14.07	6.09	3.25	Typical P-type R2R3 Myb protein (Fragment).
Os05g0139100	6.64	7.70	2.45	2.96	Basic helix-loop-helix dimerisation region bHLH domain containing protein.
Os05g0151200	6.92	4.08	4.16	3.41	(No Hit)
Os05g0151200	9.50	6.02	5.34	3.05	(No Hit)
Os05g0158600	20.18	13.77	109.22	48.13	OsGA2ox1.
Os05g0173400	2.15	5.65	2.00	2.82	(No Hit)
Os05g0174000	5.54	14.16	3.31	14.35	(No Hit)
Os05g0191500	6.13	6.54	3.10	3.07	Acid phosphatase (Class B) family protein.
Os05g0211800	17.18	20.15	28.77	32.44	(No Hit)
Os05g0211800	64.57	82.17	43.37	50.86	(No Hit)

Os05g0266800	10.35	3.57	18.95	10.18	(No Hit)
Os05g0280700	4.64	4.37	9.99	37.61	Resistance protein candidate (Fragment).
Os05g0280700	6.07	7.17	10.39	48.25	Resistance protein candidate (Fragment).
Os05g0300200	5.32	13.97	3.36	14.48	(No Hit)
Os05g0303300	2.98	2.46	6.52	11.78	(No Hit)
Os05g0304600	2.13	2.20	2.22	2.14	Linoleate:oxygen oxidoreductase (Fragment).
Os05g0304600	2.17	2.22	2.42	2.30	Linoleate:oxygen oxidoreductase (Fragment).
Os05g0326300	2.96	7.37	3.57	4.03	(No Hit)
Os05g0365200	2.93	20.11	3.61	9.70	Conserved hypothetical protein.
Os05g0372000	3.17	5.31	2.34	2.36	Conserved hypothetical protein.
Os05g0374700	306.36	9.00	210.73	5.44	Retrotransposon gag protein family protein.
Os05g0409500	3.56	2.99	2.06	19.85	MtN21 protein.
Os05g0417400	18.58	23.38	29.95	30.21	(No Hit)
Os05g0423400	2.61	6.12	3.41	7.86	PISTILLATA-like MADS box protein.
Os05g0438300	20.24	140.10	71.79	461.45	(No Hit)
Os05g0438300	23.64	170.68	65.66	432.30	(No Hit)
Os05g0465000	2.32	2.30	2.97	2.13	Conserved hypothetical protein.
Os05g0505800	2.36	3.05	2.73	2.27	(No Hit)
Os05g0516100	3.30	4.50	6.38	13.97	Conserved hypothetical protein.
Os05g0570700	14.34	26.40	13.37	30.76	(No Hit)
Os05g0580100	4.18	3.75	7.27	4.16	Hypothetical protein.
Os06g0109500	7.75	18.70	2.14	2.16	Haloacid dehalogenase-like hydrolase domain containing protein.
Os06g0127100	7.65	7.37	24.44	3.57	CBF-like protein.
Os06g0138900	4.92	2.48	3.91	2.70	Hypothetical protein.
Os06g0179500	3.13	2.47	2.18	4.67	Plant protein of unknown function family protein.
Os06g0200800	5.20	53.51	6.52	8.75	Conserved hypothetical protein.
Os06g0248300	17.56	9.35	9.08	13.66	Conserved hypothetical protein.
Os06g0261300	50.24	59.82	70.04	92.58	Hypothetical protein.
Os06g0261300	6.91	6.46	4.02	5.42	Hypothetical protein.
Os06g0265100	243.49	284.29	503.97	497.05	Hypothetical protein.
Os06g0271100	2.99	2.41	6.81	11.70	(No Hit)

Os06g0271400	564.91	536.94	88.11	125.58	Hypothetical protein.
Os06g0272900	2.84	3.74	4.56	2.87	Protein of unknown function DUF231 domain containing protein.
Os06g0278700	2.93	2.37	7.32	12.38	(No Hit)
Os06g0326900	2.70	4.19	2.63	2.03	(No Hit)
Os06g0333200	2.56	6.14	4.53	5.19	(No Hit)
Os06g0344100	3.08	2.74	7.16	12.06	(No Hit)
Os06g0366800	5.36	13.82	3.48	14.80	Conserved hypothetical protein.
Os06g0517000	4.12	3.81	10.93	10.09	Hypothetical protein.
Os06g0528800	3.07	3.31	4.50	3.60	(No Hit)
Os06g0560000	6.57	4.59	9.95	13.78	Ferroportin1 family protein.
Os06g0560000	12.29	9.21	11.72	15.80	Ferroportin1 family protein.
Os06g0572000	40.18	93.77	62.79	149.57	Hypothetical protein.
Os06g0602400	14.37	13.87	19.19	26.12	ATP-dependent RNA helicase-like protein.
Os06g0618900	3.04	2.50	6.72	12.22	(No Hit)
Os06g0671800	4.24	3.11	4.78	2.83	Cellular retinaldehyde binding/alpha-tocopherol transport family protein.
Os06g0685400	142.03	97.23	34.99	57.82	Non-protein coding transcript, uncharacterized transcript.
Os06g0696600	11.41	2.85	8.06	11.67	Xyloglucan endo-transglycosylase homolog.
Os06g0709500	3.11	2.61	7.28	12.73	(No Hit)
Os07g0110000	3.61	3.79	2.57	6.58	Hypothetical protein.
Os07g0115200	8.44	23.72	11.71	14.56	Conserved hypothetical protein.
Os07g0122000	15.16	149.23	15.48	5.30	Conserved hypothetical protein.
Os07g0123100	20.95	4.22	13.56	6.07	(No Hit)
Os07g0123200	11.33	2.79	6.61	2.30	Non-protein coding transcript, uncharacterized transcript.
Os07g0126700	2.10	4.80	3.57	4.93	(No Hit)
Os07g0152700	58.87	170.55	42.30	216.93	Glutathione S-transferase GST 20 (EC 2.5.1.18).
Os07g0153000	20.57	145.51	46.31	271.15	ZIM domain containing protein.
Os07g0154100	5.09	6.00	6.97	3.89	Viviparous-14.
Os07g0160600	5.63	2.32	2.81	6.66	Beta-Ig-H3/fasciclin domain containing protein.
Os07g0160600	6.49	2.59	2.61	8.12	Beta-Ig-H3/fasciclin domain containing protein.

Os07g0163400	13.13	26.55	12.81	32.40	(No Hit)
Os07g0195900	67.06	83.47	105.63	113.50	Polyprotein.
Os07g0200500	2.51	2.67	2.40	2.28	Conserved hypothetical protein.
Os07g0215600	2.88	2.42	7.16	12.30	(No Hit)
Os07g0229600	27.57	18.39	18.53	16.67	(No Hit)
Os07g0241500	35.98	30.52	19.58	90.90	UDP-glucuronosyl/UDP-glucosyltransferase family protein.
Os07g0241600	10.55	6.19	10.10	32.51	Conserved hypothetical protein.
Os07g0287400	10.03	32.06	18.11	51.23	Plant lipid transfer/seed storage/trypsin-alpha amylase inhibitor domain containing protein.
Os07g0291400	2.97	8.84	2.19	2.94	Conserved hypothetical protein.
Os07g0297500	18.70	8.00	7.30	7.36	Actin-binding, cofilin/tropomyosin type domain containing protein.
Os07g0429600	11.51	13.11	38.07	117.62	Conserved hypothetical protein.
Os07g0429700	6.92	6.12	20.11	63.09	Conserved hypothetical protein.
Os07g0431000	5.48	4.97	16.13	37.55	En/Spm-like transposon proteins family protein.
Os07g0442800	13.55	46.75	170.49	24.62	Conserved hypothetical protein.
Os07g0446800	6.71	15.81	9.46	20.12	Hexokinase.
Os07g0457200	15.15	18.54	10.40	12.39	Non-protein coding transcript, putative npRNA.
Os07g0471900	2.99	2.67	2.56	3.75	Basic helix-loop-helix dimerisation region bHLH domain containing protein.
Os07g0488800	3.10	2.58	7.08	11.79	(No Hit)
Os07g0511400	65.94	36.05	4.12	12.60	Hypothetical protein.
Os07g0536500	3.00	2.50	7.06	12.57	(No Hit)
Os07g0538000	6.79	34.53	29.75	5.70	Beta-1,3-glucanase precursor.
Os07g0611100	3.04	2.44	6.97	12.20	(No Hit)
Os07g0620600	3.79	5.51	3.89	2.18	Conserved hypothetical protein.
Os07g0642200	8.07	8.66	4.52	18.60	Lipolytic enzyme, G-D-S-L family protein.
Os07g0648500	7.74	69.36	19.41	132.45	(No Hit)
Os07g0653100	4.17	5.56	4.96	3.65	(No Hit)
Os07g0653100	4.20	5.63	4.96	3.67	(No Hit)
Os07g0655800	9.46	28.14	14.42	57.73	Conserved hypothetical protein.
Os07g0657100	7.71	4.08	3.13	2.68	Glyoxalase/bleomycin resistance protein/dioxygenase domain containing protein.

Os07g0694200	310.83	916.69	556.00	1383.12	Conserved hypothetical protein.
Os08g0106900	2.04	3.77	2.92	4.89	Transposase, IS4 domain containing protein.
Os08g0124500	2.46	5.46	2.31	2.49	Resistance protein candidate (Fragment).
Os08g0124600	13.56	27.32	12.41	31.42	(No Hit)
Os08g0140700	23.32	38.31	28.70	79.34	Conserved hypothetical protein.
Os08g0149600	2.94	2.45	7.22	12.43	(No Hit)
Os08g0178900	11.06	994.58	40.85	1092.26	Homeodomain-like containing protein.
Os08g0184200	5.60	14.05	3.46	14.65	(No Hit)
Os08g0213100	11.83	13.56	10.22	6.24	(No Hit)
Os08g0239900	10.70	15.70	29.96	79.10	Sulfotransferase family protein.
Os08g0240000	9.61	12.60	9.87	27.17	STF-1 (Fragment).
Os08g0240000	10.10	13.64	9.61	28.92	STF-1 (Fragment).
Os08g0240200	7.85	24.55	12.32	29.37	Hypothetical protein.
Os08g0246800	5.80	5.50	6.97	5.52	Hypothetical protein.
Os08g0246800	2.57	2.86	3.30	2.87	Hypothetical protein.
Os08g0274300	5.48	13.76	3.45	14.38	(No Hit)
Os08g0289400	13.10	27.00	13.88	34.71	Zn-finger, CCHC type domain containing protein.
Os08g0290700	4.70	3.62	2.96	2.99	O-methyltransferase, family 2 domain containing protein.
Os08g0299900	3.99	4.18	4.65	5.87	Cyclin-like F-box domain containing protein.
Os08g0311500	5.51	8.27	5.02	5.95	(No Hit)
Os08g0316400	27.97	17.94	18.91	16.71	(No Hit)
Os08g0325600	5.33	14.33	3.58	15.26	(No Hit)
Os08g0327900	2.96	2.44	6.81	11.78	(No Hit)
Os08g0344700	6.02	55.68	13.47	32.96	Retrotransposon gag protein family protein.
Os08g0348800	3.00	2.44	6.87	11.59	(No Hit)
Os08g0355400	2.66	4.97	3.68	14.49	Plant protein of unknown function family protein.
Os08g0355400	2.42	3.62	3.68	10.61	Plant protein of unknown function family protein.
Os08g0360700	4.90	11.71	5.42	30.69	Conserved hypothetical protein.
Os08g0371200	8.68	24.26	5.21	5.47	Conserved hypothetical protein.
Os08g0389500	12.56	21.24	26.10	9.96	Viral coat and capsid protein family protein.

Os08g0408300	14.25	19.20	4.99	12.42	Conserved hypothetical protein.
Os08g0410900	107.03	3.72	143.03	14.27	Non-protein coding transcript, uncharacterized transcript.
Os08g0428400	2.07	2.26	2.78	2.52	ZIM domain containing protein.
Os08g0428800	8.07	17.17	38.85	27.54	High mobility group proteins HMG-I and HMG-Y family protein.
Os08g0474000	4.76	8.98	34.49	2.01	AP2 domain containing protein RAP2.6 (Fragment).
Os08g0484200	2.50	4.76	2.46	2.45	Zn-finger, RING domain containing protein.
Os08g0487300	2.60	6.09	4.50	5.20	Zn-finger, CCHC type domain containing protein.
Os08g0509400	2.91	2.38	5.55	12.68	Glycoside hydrolase, family 1 protein.
Os08g0509400	3.29	2.33	4.03	10.23	Glycoside hydrolase, family 1 protein.
Os08g0518900	4.66	12.21	5.24	8.12	Glycoside hydrolase, family 18 protein.
Os08g0518900	3.60	8.56	3.06	5.07	Glycoside hydrolase, family 18 protein.
Os09g0122500	3.25	10.21	2.28	2.89	Hypothetical protein.
Os09g0244600	4.12	110.21	12.39	83.20	Conserved hypothetical protein.
Os09g0272400	3.11	8.03	3.49	4.05	(No Hit)
Os09g0275400	7.15	18.14	44.57	22.81	Cytochrome P450 family protein.
Os09g0292900	24.44	6.46	26.29	55.38	Cyclin-like F-box domain containing protein.
Os09g0319800	4.35	18.39	9.17	2.69	Terpenoid cylases/protein prenyltransferase alpha-alpha toroid domain containing protein.
Os09g0392400	3.10	4.35	2.67	2.37	PDR13 ABC transporter.
Os09g0463800	2056.08	1484.29	2147.91	2729.73	Conserved hypothetical protein.
Os09g0488100	11.37	3.00	17.33	3.02	(No Hit)
Os09g0491100	12.27	8.67	7.22	6.16	Beta-primeverosidase (EC 3.2.1.149).
Os09g0491100	10.45	8.52	6.83	8.13	Beta-primeverosidase (EC 3.2.1.149).
Os09g0491100	8.62	5.85	9.14	8.53	Beta-primeverosidase (EC 3.2.1.149).
Os09g0524300	9.65	5.23	9.13	13.69	Multi antimicrobial extrusion protein MatE family protein.

Os09g0554200	3.60	2.37	4.80	3.93	Zn-finger, RING domain containing protein.
Os10g0107100	3.48	12.38	6.91	10.11	(No Hit)
Os10g0158900	5.48	13.88	3.39	14.48	(No Hit)
Os10g0162200	5.12	10.59	4.05	3.40	Zn-finger, CCHC type domain containing protein.
Os10g0169200	3.11	16.59	8.27	23.04	Conserved hypothetical protein.
Os10g0320100	4.52	6.48	6.69	12.98	Flavonoid 3'-monooxygenase (EC 1.14.13.21) (Flavonoid 3'-hydroxylase) (Cytochrome P450 75B2).
Os10g0334100	3.07	2.57	6.86	12.09	(No Hit)
Os10g0343500	2.17	2.58	3.01	6.07	(No Hit)
Os10g0353100	2.82	24.83	6.75	8.55	Conserved hypothetical protein.
Os10g0371600	1428.36	4067.43	2204.12	5897.12	Conserved hypothetical protein.
Os10g0419400	36.44	53.55	11.93	19.22	Submergence induced protein 2.
Os10g0419400	36.12	53.33	11.48	18.77	Submergence induced protein 2.
Os10g0419400	33.62	49.12	10.86	17.01	Submergence induced protein 2.
Os10g0444300	2.38	6.31	2.01	2.97	(No Hit)
Os10g0452100	26.50	37.82	3.60	129.71	Glycine-rich protein (Fragment).
Os10g0505900	3.62	4.79	9.67	6.31	Conserved hypothetical protein.
Os10g0520000	3.25	7.32	4.41	8.05	Spectrin repeat containing protein.
Os10g0546100	139.58	238.60	93.30	822.11	Pollen Ole e 1 allergen and extensin domain containing protein.
Os10g0565200	2.21	4.54	2.36	2.04	Beta-amylase PCT-BMYI (EC 3.2.1.2).
Os10g0567900	50.49	48.85	51.79	68.88	F-box protein interaction domain containing protein.
Os10g0578600	3.33	2.84	2.97	5.04	Conserved hypothetical protein.
Os11g0125900	3.52	5.63	11.23	2.38	Nucleoside phosphatase GDA1/CD39 family protein.
Os11g0151400	14.25	16.39	31.97	3.19	Cytochrome P450 family protein.
Os11g0181200	4.75	10.19	4.56	10.24	Hypothetical protein.
Os11g0192600	2.95	2.51	7.53	12.81	(No Hit)
Os11g0208700	9.17	6.49	4.99	2.13	Protein kinase family protein.
Os11g0212900	10.61	8.51	17.49	4.31	Hypothetical protein.
Os11g0213000	22.77	19.18	10.54	7.32	Protein kinase domain containing protein.
Os11g0230900	2.95	2.35	6.54	11.49	(No Hit)

Os11g0235000	2.88	2.38	6.77	11.75	(No Hit)
Os11g0260000	4.10	7.36	2.01	2.32	Hypothetical protein.
Os11g0274100	9.00	6.42	6.97	12.59	Protein kinase domain containing protein.
Os11g0278900	9.85	3.86	17.23	3.07	(No Hit)
Os11g0278900	10.54	4.13	20.94	5.56	(No Hit)
Os11g0280900	14.15	37.60	22.96	45.30	Conserved hypothetical protein.
Os11g0282700	46.93	42.75	57.28	54.34	Homeodomain-like containing protein.
Os11g0303800	3.70	4.08	3.12	2.76	Homeodomain-like containing protein.
Os11g0418900	8.90	63.15	4.38	5.94	Conserved hypothetical protein.
Os11g0457300	22.38	25.78	43.19	8.10	Flavonol 3-O-glucosyltransferase (EC 2.4.1.91) (UDP-glucose flavonoid 3-O-glucosyltransferase) (Anthocyanin rhamnosyl transferase).
Os11g0471200	117.86	252.12	111.61	230.89	Conserved hypothetical protein.
Os11g0477400	6.03	15.64	4.97	2.64	Transposase (Fragment).
Os11g0491900	6.90	3.45	28.03	19.95	(No Hit)
Os11g0512000	4.69	12.98	8.09	40.68	No apical meristem (NAM) protein domain containing protein.
Os11g0519100	83.20	556.71	83.36	813.06	Conserved hypothetical protein.
Os11g0540600	6.10	10.36	4.53	4.26	Plant protein of unknown function family protein.
Os11g0550300	2.80	3.53	2.67	2.11	Conserved hypothetical protein.
Os11g0557000	2.65	2.60	3.57	4.11	Phytosulfokine family protein.
Os11g0595900	44.52	169.85	90.29	215.37	Hypothetical protein.
Os11g0624700	8.20	14.40	2.53	2.67	(No Hit)
Os11g0625000	4.38	3.07	4.46	3.64	(No Hit)
Os11g0626700	2.06	3.22	3.97	3.31	Hypothetical protein.
Os11g0632200	3.28	2.56	2.72	3.40	Conserved hypothetical protein.
Os11g0636200	17.71	120.04	42.70	260.89	(No Hit)
Os11g0674500	2.66	3.48	4.46	5.32	NBS-LRR-like protein D.
Os11g0681700	7.10	5.41	3.51	10.09	Hypothetical protein.
Os11g0692300	3.57	4.10	2.42	2.36	Bacterial blight resistance protein.
Os12g0113600	3.91	3.35	2.91	4.07	Hypothetical protein.
Os12g0117700	5.56	21.95	4.50	9.51	Hypothetical protein.
Os12g0130300	5.97	8.34	7.24	2.82	Resistance protein candidate (Fragment).

Os12g0156600	3.03	7.74	3.40	4.27	(No Hit)
Os12g0186600	6.50	17.45	3.87	16.35	Conserved hypothetical protein.
Os12g0202700	4.68	4.42	3.13	2.70	O-methyltransferase, family 2 protein.
Os12g0218900	19.56	45.98	16.49	35.21	(No Hit)
Os12g0222800	15.62	32.73	4.86	8.09	(No Hit)
Os12g0229700	3.03	2.48	6.74	11.56	(No Hit)
Os12g0244700	18.00	62.30	14.83	70.91	Non-protein coding transcript, uncharacterized transcript.
Os12g0256000	3.61	2.73	3.28	3.06	Esterase/lipase/thioesterase domain containing protein.
Os12g0270900	2.84	2.98	2.46	3.44	Sulfotransferase family protein.
Os12g0275100	45.80	81.53	32.57	98.37	Hypothetical protein.
Os12g0275200	39.03	269.55	63.73	542.89	Conserved hypothetical protein.
Os12g0408000	14.55	17.23	11.50	34.00	Protein of unknown function DUF594 family protein.
Os12g0415400	12.05	38.82	47.07	137.40	Histone H3.
Os12g0415800	32.23	100.31	84.52	270.54	Histone H3.2 (Fragment).
Os12g0428300	39.55	36.97	70.36	80.47	Retrotransposon gag protein family protein.
Os12g0440400	11.44	210.08	22.36	225.37	Hypothetical protein.
Os12g0463300	13.37	27.00	12.10	29.59	(No Hit)
Os12g0467500	8.18	26.21	16.59	21.57	SWIM Zn-finger domain containing protein.
Os12g0473100	2.93	2.37	6.73	11.39	(No Hit)
Os12g0483500	2.36	6.75	3.13	10.70	(No Hit)
Os12g0491400	19.22	20.32	20.85	13.76	(No Hit)
Os12g0501200	2.03	5.59	2.05	3.15	(No Hit)
Os12g0516800	10.33	12.39	3.35	3.35	Protein of unknown function DUF231 domain containing protein.
Os12g0582700	6.85	3.48	9.68	3.69	Cytochrome P450 family protein.
Os12g0582700	41.75	18.63	9.85	6.65	Cytochrome P450 family protein.
Os12g0593700	3.16	2.82	7.50	13.73	(No Hit)
Os12g0636500	7.92	6.48	3.61	2.82	Hypothetical protein.
Os12g0637100	2.26	3.29	2.60	2.22	Purple acid phosphatase (EC 3.1.3.2).
osa-miR419|Os12	16.35	117.78	7.80	189.39	miRNA

* Probe name is described as RAP database (http://rapdb.dna.affrc.go.jp/).

** Gene description is described as RAP database,

*** Genes annotated as TEs bt RAP database or MSU Rice Genome Annotation Release 7 (http://rice.plantbiology.msu.edu) database is highlighted.

mut; *Osddm1s* mutant, N; Nipponbare.

Table 3. Down-regulated genes in *Osddm1s* mutant seedlings.

ProbeName	mut1/N1 Ratio	mut2/N2 Ratio	mut1-2/N2 Ratio	mut2-2/N2 Ratio	Description
Os01g0198500	0.40	0.27	0.34	0.16	Conserved hypothetical protein.
Os01g0304400	0.31	0.50	0.30	0.29	Conserved hypothetical protein.
Os01g0307500	0.48	0.43	0.44	0.32	Cation transporter family protein.
Os01g0550800	0.16	0.02	0.21	0.26	Protein of unknown function DUF239 domain containing protein.
Os01g0974600	0.30	0.26	0.20	0.37	RNA-binding region RNP-1 (RNA recognition motif) domain containing protein.
Os02g0221900	0.25	0.39	0.25	0.32	Cytochrome P450 family protein.
Os02g0221900	0.20	0.43	0.33	0.35	Cytochrome P450 family protein.
Os02g0481900	0.35	0.02	0.24	0.04	Putative 5-3 exonuclease domain containing protein.
Os02g0504800	0.27	0.10	0.29	0.13	Acyl-[acyl-carrier-protein] desaturase, chloroplast precursor (EC 1.14.19.2) (Stearoyl-ACP desaturase).
Os02g0587300	0.45	0.39	0.44	0.35	Hypothetical protein.
Os02g0731900	0.39	0.36	0.45	0.41	Naringenin-chalcone synthase family protein.
Os03g0165900	0.47	0.09	0.20	0.31	MS5-like protein (Fragment).
Os03g0226200	0.37	0.25	0.10	0.37	Non-symbiotic hemoglobin 2 (rHb2) (ORYsa GLB1b).
Os03g0226200	0.38	0.25	0.10	0.34	Non-symbiotic hemoglobin 2 (rHb2) (ORYsa GLB1b).
Os03g0609500	0.50	0.29	0.12	0.28	Protein of unknown function DUF260 domain containing protein.
Os03g0745000	0.29	0.37	0.24	0.45	Heat shock factor (HSF)-type, DNA-binding domain containing protein.
Os04g0352400	0.46	0.43	0.38	0.27	Immunophilin FK506 binding protein FKBP12.
Os04g0463700	0.47	0.45	0.34	0.40	General negative regulator of transcription subunit 4.
Os04g0526800	0.43	0.31	0.27	0.41	GRAM domain containing protein.

Os05g0107900	0.43	0.40	0.27	0.35	Drought induced 19 family protein.
Os05g0202800	0.32	0.28	0.30	0.23	Plant metallothionein, family 15 protein.
Os05g0207200	0.44	0.44	0.41	0.34	Protein prenyltransferase domain containing protein.
Os05g0397300	0.41	0.10	0.46	0.29	Conserved hypothetical protein.
Os05g0511500	0.44	0.35	0.37	0.45	Lipoic acid synthetase, mitochondrial precursor (Lip-syn) (Lipoate synthase).
Os05g0547600	0.26	0.30	0.39	0.30	Conserved hypothetical protein.
Os05g0595100	0.21	0.22	0.45	0.32	UDP-glucose 4-epimerase.
Os06g0108500	0.29	0.23	0.46	0.34	(No Hit)
Os06g0206700	0.40	0.49	0.47	0.48	Kinesin, motor region domain containing protein.
Os06g0229000	0.33	0.08	0.35	0.23	FtsH protease (VAR2) (Zinc dependent protease).
Os06g0329300	0.36	0.19	0.17	0.13	Conserved hypothetical protein.
Os06g0495100	0.27	0.06	0.16	0.11	Conserved hypothetical protein.
Os06g0554200	0.43	0.26	0.30	0.29	Conserved hypothetical protein.
Os06g0624200	0.42	0.23	0.41	0.35	Non-protein coding transcript, uncharacterized transcript.
Os07g0611700	0.37	0.13	0.05	0.44	Conserved hypothetical protein.
Os07g0645200	0.21	0.12	0.22	0.15	Hypothetical protein.
Os08g0286500	0.36	0.44	0.15	0.32	Non-protein coding transcript, uncharacterized transcript.
Os09g0272300	0.46	0.43	0.31	0.42	Glycoside hydrolase, family 17 protein.
Os09g0388400	0.22	0.26	0.42	0.26	Cof protein family protein.
Os09g0520800	0.38	0.28	0.35	0.41	Galactose-binding like domain containing protein.
Os09g0546900	0.49	0.43	0.23	0.33	Auxin induced protein.
Os10g0154300	0.37	0.35	0.35	0.39	(No Hit)
Os10g0204400	0.28	0.29	0.19	0.30	Phosphoenolpyruvate carboxykinase.
Os10g0414700	0.39	0.48	0.38	0.29	KH, type 1 domain containing protein.
Os10g0425700	0.50	0.50	0.43	0.43	MATH domain containing protein.
Os10g0478100	0.48	0.28	0.23	0.19	Heat shock protein DnaJ,

Os10g0478100	0.48	0.28	0.23	0.19	N-terminal domain containing protein.
Os10g0502200	0.46	0.28	0.36	0.25	Non-protein coding transcript, unclassifiable transcript.
Os10g0563300	0.42	0.48	0.44	0.44	WD40-like domain containing protein.
Os11g0220100	0.32	0.22	0.35	0.48	Hypothetical protein.
Os11g0461000	0.36	0.10	0.32	0.33	Peptidase S10, serine carboxypeptidase family protein.
Os12g0406000	0.26	0.26	0.25	0.45	Hypothetical protein.
Os12g0406000	0.27	0.23	0.26	0.44	Hypothetical protein.
Os12g0523000	0.35	0.22	0.28	0.20	(No Hit)
Os12g0595600	0.27	0.29	0.42	0.40	(No Hit)

DISCUSSION

Using HR-promoted knock-out GT, an identical mutation for each of the rice *DDM1a* and *DDM1b* genes encoding putative SWI2/SNF2-like chromatin remodeling ATPases were reproducibly generated (Figure 1). No effect of *OsDDM1a* disruption on cytosine methylation was observed at any repetitive sequences investigated (Figure 5). In contrast, *OsDDM1b* disruption caused decreased cytosine methylation in repetitive sequences (Figure 6). Subsequently generated *OsDDM1a* and *OsDDM1b* double disruptant mutants exhibited drastic reduction of cytosine methylation. Double disruption derepressed the transcription of several TEs suggesting an essential role for rice DDM1s on epigenetic regulation of repetitive sequences.

OsDDM1 gene is a two-copy gene in rice. The exon-intron structures and amino acid sequences are highly conserved between *OsDDM1a* and *OsDDM1b*, suggesting that one of these might have arisen from the other by gene duplication. The gene expression patterns are slightly different each other; the transcripts of *OsDDM1b* accumulate more abundantly than those of *OsDDM1a* in all the tissues examined except in anther (Figure 2). It was reported that *Brassica rapa* also has two copies of *DDM1* gene, *BrDDM1a* and *BrDDM1b*, which show high conservation in their gene structures and amino acid sequences (Fujimoto *et al.*, 2008). The study of knock-down plants with reduced levels of

BrDDM1a/BrDDM1b expression could not distinguish the effect of each BrDDM1a and BrDDM1b. This is because the knock-down plants were generated by RNAi procedure expressing a double-strand RNA which could down-regulate both *BrDDM1a* and *BrDDM1b* transcripts at the same time (Fujimoto *et al.*, 2008). In the present study, by generating HR-promoted knock-out-targeted mutants for each of *OsDDM1a* and *OsDDM1b*, I found the different degree of contribution of OsDDM1a and OsDDM1b on cytosine methylation in rice. OsDDM1b, which is more expressed than OsDDM1a, plays more essential roles than OsDDM1a. *OsDDM1a* single disruptant mutants exhibited no obvious effect on cytosine methylation, however, *OsDDM1a* and *OsDDM1b* double disruptant mutants showed additive effects comparing to *OsDDM1b* single disruptant mutant, suggesting OsDDM1a has some redundant function with OsDDM1b at certain extent. Although I could not detect any significant demethylation effect in *OsDDM1a* mutants at least in the T_2 generation, it would be interesting to ask if *OsDDM1a* has some specific function in anther where *OsDDM1a* transcripts exceptionally accumulate more abundantly than those of *OsDDM1b*. During male gametogenesis, tricellular pollen comprising a vegetative nucleus (VN) and two sperm cells (SCs) is produced in anther. The SCs that have methylated DNA provide genetic information to subsequent generations through double fertilization, and the VN that has reduced DNA methylation in Arabidopsis supports the SCs

before fertilization. Recent studies reported that the DNA methylation in SCs and VN appeared to be regulated by DDM1 in part, and resulting transient reactivation of TEs and the generations of small interfering RNAs in VN may ensure the genomic integrity of SCs be silencing TEs in the SCs in Arabidopsis (Slotkin *et al.*, 2009; Law and Jacobsen, 2010). If such DNA demethylation occurs in the rice vegetative nucleus, it will be intriguing to ask whether DDM1a, not DDM1b, has specific function in that process

The *OsDDM1a* and *OsDDM1b* double disruptant mutants exhibited the severe reduction in cytosine methylation at all the examined repetitive sequences, implying genome-wide demethylation in the mutants. In the double diruptant mutants, the transcription of several TEs including retrotransposon *RIRE7* and *Mutator-like* DNA transposon, was also derepressed, suggesting an essential role for DDM1s on epigenetic regulation of repetitive sequences in rice genome. I could detect demethylation in *RIRE7* in *OsDDM1b* disruptant, however, the transcription level of *RIRE7* in this single mutant was more or less similar to sibling plants with WT genotype. The derepression of *RIRE7* transcription appears to be correlated with the methylation level in *RIRE7*. The reduction of cytosine methylation in *OsDDM1b* mutant may not be enough to reactivate the *RIRE7* transcription. In Arabidopsis *ddm1* mutants, in parallel with demethylation, reactivation and transposition of endogenous TEs was observed causing TEs

insertion and/or ectopic gene expression that resulted in specific phenotype. I could observe the immediate hypomethylation occurring in repetitive sequences in double disruptant mutants, however, I could not detect any obvious phenotype. Unlike most other plant species such as rice and *B. rapa*, only a small part of the genome in Arabidopsis consists of TEs, and most of the endogenous TE families are present in only a few copies in the genome. It was reported that in *B. rapa* developmental abnormalities were not observed in the T_0, T_1, T_2, and F_1 *BrDDM1a/BrDDM1b* RNAi lines, suggesting that the effect of *ddm1* background on developmental abnormalities may not be simply correlated with quantity of repetitive sequences (Fujimoto *et al.*, 2008). Recent study of the transposon *mPing* massive amplification in rice revealed that the vast majority of *mPing* insertions caused modest impact on gene transcription reflecting avoidance of exon insertions by *mPing* (Naito *et al.*, 2009). This might be one of the ways for a host genome to survive a rapid burst of TE insertions and it might be a reason why I could not detect obvious phenotype in double disruptant mutants. The study also indicated that a subset of *mPing* insertion can render adjacent genes stress inducible (Naito *et al.*, 2009). It would be possible, alternatively, that I could detect some phenotypes of rice *DDM1s* double disruptant mutants under certain stress conditions.

REFERENCES

Brzeski, J. and Jerzmanowski, A. (2003) *Deficient in DNA Methylation 1 (DDM1)* defines a novel family of chromatin-remodeling factors. *J. Biol. Chem.* **278**, 823-828.

Chan, S.W.-L., Henderson, I.R. and Jacobsen, S.E. (2005) Gardening the genome; DNA methylation in *Arabidopsis thaliana*. *Nat. Rev. Genet.* **6**, 351-360.

De la Fuente, R., Baumann, C., Fan, T., Schmidtmann, A., Dobrinski, A. and Muegge, K. (2006) Lsh is required for meiotic chromosome synapsis and retrotranspososn silencing in female germ cells. *Nat. Cell. Biol.* **8**, 1448-1454.

Dennis, K., Fan, T., Geiman, T., Yan, Q. and Muegge, K. (2001) Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.* **15**, 2940-2944.

Fujimoto, R., Sasaki, T., Inoue, H. and Nishio, T. (2008) Hypermethylation and transcriptional reactivation of retrotransposon-like sequences in *ddm1* transgenic plants of *Brassica rapa*. *Plant Mol. Biol.* **66**, 463-473.

Gendrel, A.V., Lippman, Z., Yordan, C., Colot, V. and Martienssen, R.A. (2002) Dependance of heterochromatic histone H3 methylation pattern on the *Arabidopsis* gene *DDM1*. *Science* **297**, 1871-1873.

Goff, S.A., Ricke, D., Lan, T.-H., Presting, G., Wang, R., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., *et al.* (2002) A draft sequence of the rice genome (*Osyza sativa* L. ssp. *japonica*). Science **296**, 92-100.

Jeddeloh, J.A., Stokes, T.L. and Richards, E.J. (1999) Maintenance of genomic methylation requires a SWI2/SNF2-loke protein. *Nat. Genet.* **22**, 94-97.

Johzuka-Hisatomi, Y., Terada, R. and Iida, S. (2008) Efficient transfer of base changes from a vector to the rice genome by homologous recombination: involvement of heteroduplex formation and mismatch correction. *Nucleic Acids Res.* **36**, 4727-4735.

Kakutani, T., Jeddeloh, J.A., Flowers, S.K., Munakata, K. and Richards, E.J. (1996) Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc. Natl. Acad. Sci. USA* **93**, 12406-12411.

Kumekawa, N., Ohmido, N., Fukui, K., Ohtsubo, E., and Ohtsubo, E. (2001) A new *gypsy*-type retrotransposon, *RIRE7*: preferential insertion into the tandem repeat sequence TrsD in pericentromeric heterochromatin regions of rice chromosomes. *Mol. Genet. Genomics* **265**, 480-488.

Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204-220.

Lippman, Z., May, B., Yordan, C., Singer, T. and Martienssen, R. (2003) Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PloS Biol.* **1**, E67.

Lippman, Z., Gendrel, A.V., Black, M., Vaughn, M.W., Dedhia, N., McCombie, W.R., Lavine, K., Mittal, V., May, B., Kasschau, K.D., *et al.* (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* **430**, 471-476.

Muegge, K. (2005) Lsh, a guardian of heterochromatin at repeat elements. *Biochem. Cell. Biol.* **83**, 548-554.

Naito, K., Zhang, F., Tsukyama, T., Saito, H., Hancock, C.N., Richardson, A.O., Okumoto, Y., Tanisaka, T. and Wessler, S.R. (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*, **461**, 1130-1134.

Rajhi, I., Yamauchi, T., Takahashi, H., Nishiuchi, S., Shiono, K., Watanabe, R., Miki A., Nagamura, Y., Tsutsumi, N, Nishizawa, K.K. and Nakazono M. (2011) Identification of genes expressed in maize root cortical cells during

lysigenous aerenchyma formation using laser microdissection and microarray analyses. *New Phytologist* **190**, 351-368.

Slotkin, R.K., Vaughn, M., Bonges, F., Tanurdzic, M., Becker, J.D., Feijó, J.A. and Martienssen, R.A. (2009) Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. *Cell* **136**, 461-472.

Takeda, S., Sugumoto, K., Otsuki, H. and Hirochika, H. (1999) A 13-bp *cis*-regulatory element in the LTR promoter of the tobacco retrotransposon *Tto1* is involved in responsiveness to tissue culture, wounding, methyla jasmonate and fungal elicitors. Plant J **18,** 383-393.

Terada, R., Urawa, H., Inagaki, Y., Tsugane, K. and Iida, S. (2002) Efficient gene targeting by homologous recombination in rice. *Nat. Biotechnol.* **20**, 1030-1034.

Terada, R., Johzuka-Hisatomi, Y., Saitoh, M., Asao, H. and Iida, S. (2007) Gene targeting by homologous recombination as a biotechnological tool for rice functional genomics. *Plant Physiol.* **144**, 846-856.

Vongs, A., Kakutani, T., Martienssen R.A. and Richards, E.J. (1993) *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**, 1926-1928.

Yamauchi, T., Johzuka-Hisatomi, Y., Fukada-Tanaka, S., Terada, R., Nakamura, I. and Iida, S. (2009) Homologous recombination-mediated knock-in targeting of the *MET1a* gene for a maintenance DNA

methyltransferase reproducibly reveals dosage-dependent spatiotemporal gene expression in rice. Plant J. **60**, 386-396.

Yu, J., Hu, S., Wang, J., Wong, G.K.-S., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science* **296**, 79-92.

Zhu, J.-K. (2009) Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43**, 143-166.

General Discussion

Rice seeds consist of the diploid embryo, which has root and shoot apical meristems, and the triploid endosperm, which support energy consumption around the world. The shoot apical meristem produces leaves and bisexual flowers which have both male and female reproductive structures (anther and gynocium, respectively) enclosed by a single set of glumes (lemma and palea). The haploid phase begins as each plant produces both micro- and mega-gametophytes in florets. In anthers, following meiosis, each haploid cell undergoes mitotic divisions that result in pollen grains containing three genetically identical nuclei, two sperm nuclei and one vegetative nucleus, in each. When a pollen grain lands on pistil, one haploid sperm cell fertilizes a single haploid egg to produce a diploid embryo. In parallel, the second haploid sperm cell fertilizes two female nuclei to produce a triploid endosperm. In these processes of the rice life cycle, accurate regulation of gene expression is crucial for defining cellular identities and coordinating developmental programs. In plants, unlike animals, germ line is not set apart from soma early in its development. In late developmental stage, floral transition is promoted, and meristematic cells in the shoot apex switch from producing vegetative organs to producing the reproductive organs. This means that when a mutation occurs in a subset of shoot apical

meristem cells during its vegetative growth under various environmental conditions, the mutation can be inherited to its progeny and can affect on genome function in the next generations. Recent surveys of genome sequences of many organisms including rice have revealed that most genomes, which transmit genetic and epigenetic information, consist of abundant transposon-related sequences including RNA-based mobile elements such as retrotransposons and retroviruses, and DNA transposons. They have contributed to large-scale genome insertions, rearrangements and allelic diversification during evolution. In addition, they are potentially mutable to the organisms because their repeated units can lead to spurious homologous recombination, and their ability to transpose can cause disruption or misregulation of important genes suggesting host organisms require the systems to fight against the potential hazard.

The importance of keeping a proper state of epigenetic modifications during the life cycle has most drastically been shown by studying organisms' mutants in their modification activities, and it is one of the missing parts in epigenetic study in rice. In this thesis, using HR-promoted GT, identical mutations for the rice *ROS1a* gene encoding a putative cytosine DNA demethylase and for the rice *DDM1a/DDM1b* genes encoding SWI2/SNF2 chromatin remodeling factors were reproducibly generated. Obtained knock-in/knock-out mutants have shown that proper cytosine demethylation is indispensable in order to facilitate

normal developmental processes in rice life cycle, such as fertilization of male gametophyte and endosperm development, and also that proper cytosine methylation is necessary for protection of rice genome from undesired events, such as potentially harmful reactivation of TEs. These results indicate that in rice genome, DNA methylation, which is established as the balance between DNA methylation and demethylation activities, can play roles to regulate accurate gene expression for the developmental program and to keep a proper state of the genome during the life cycle (Figure 1).

One question to be answered is to identify the genes under ROS1a regulation. Although I could not demonstrate direct gene regulation via ROS1a-promoted DNA demethylation, ROS1a is postulated to activate the gene expression through DNA demethylation for, at least, normal male gametophyte fertilization and normal endosperm development. Identification of such genetic components and their detailed functional elucidation will be the key to understanding the molecular mechanisms of these processes. Because these processes directly connected to the seeds production of rice plants, it would also provide important information for agribiosciences.

Another key issue would be searching for rice epi-alleles in *Osddm1a Osddm1b* double mutants. Although I could detect demethylation in repetitive sequences, obvious phenotypic changes were not observed in *Osddm1a*

Osddm1b double mutants compared with WT siblings. Plant species other than Arabidopsis have their own unique genetic traits controlled by epigenetic gene regulation, such as self-incompatible system in *Brassicaceae* (reviewed in Shiba and Takayama, 2012) and tomato fruit ripening process (reviewed in Seymour *et al.*, 2008). Hypomethylated rice mutants would be unique sources to search epigenetically regulated traits that could be new resources for rice breeding. In Arabidopsis, *ddm1* mutants show only slight morphological change at the first generation and notable morphological abnormalities have been observed after repeated self-pollination for several generations (Vongs *et al.*, 1993; Kakutani *et al.*, 1996). To understand molecular mechanism further in rice genome and to generate genetic resources for breeding, it would be interesting to do large scale screening for epi-alleles in the *Osddm1a Osddm1b* double mutants, which are viable and fertile enough to propagate next generations, after several self-pollination or under some stress conditions.

In summary, these studies would provide valuable information on DNA methylation/demethylation-mediated epigenetic rice genome regulation and unique tools for further studies in this area not only for basic science of rice but also for applied crop research.

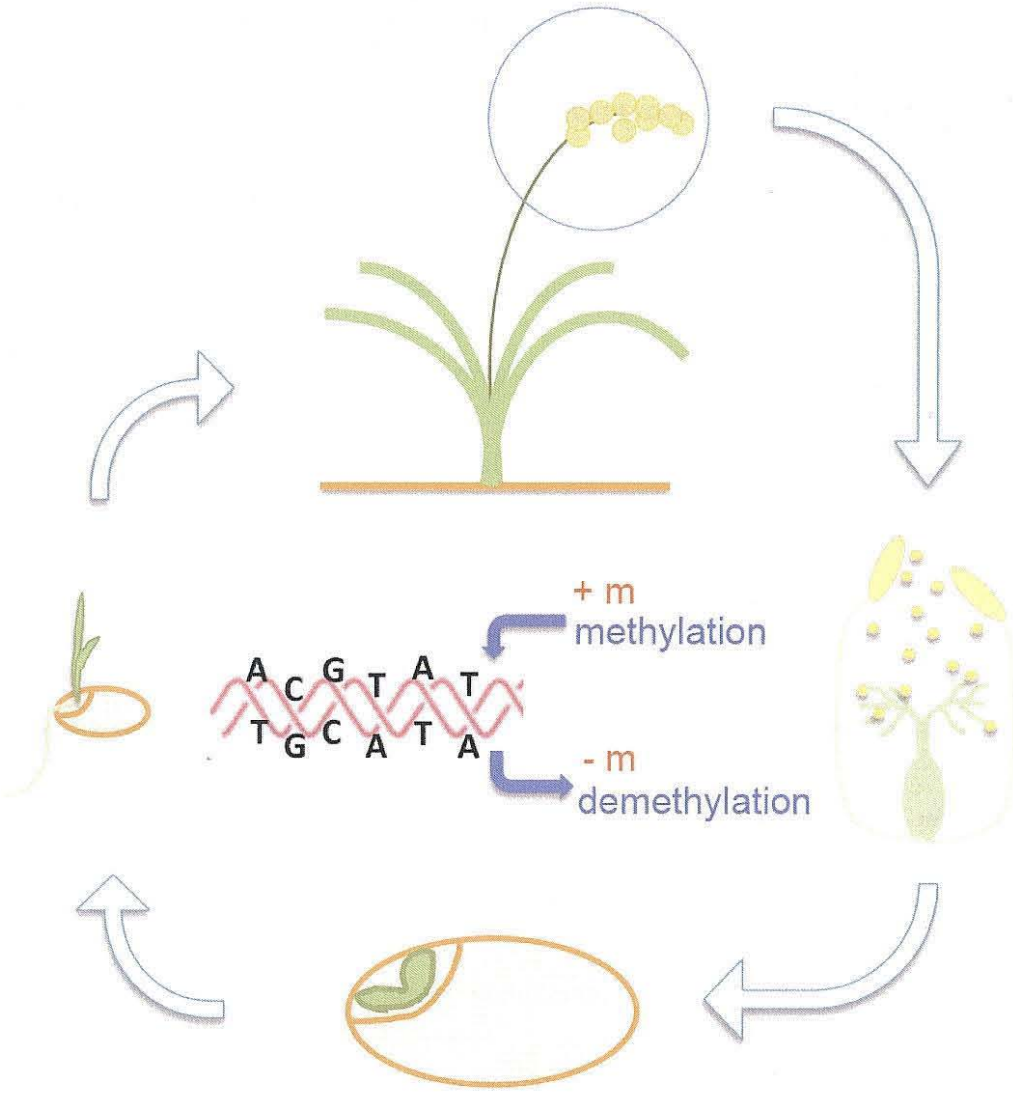


Figure 1

Figure 1. Diagram of rice life cycle.

DNA methylation/demethylation play important roles in development and genome protection during rice life cycle.

REFERENCES

Kakutani, T., Jeddeloh, J.A., Flowers, S.K., Munakata, K. and Richards, E.J. (1996) Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc. Natl. Acad. Sci. USA* **93**, 12406-12411.

Seymour G., Poole, M., Manning, K. and King, G.J. (2008) Genetics and epigenetics of fruit development and ripening. *Curr. Opin. Plant Biol.* **11.** 58-63.

Shiba, H. and Takayama, S. (2012) Epigenetic regulation of monoallelic gene expression. *Develop. Growth Differ.* **54.** 120-128.

Vongs, A., Kakutani, T., Martienssen R.A. and Richards, E.J. (1993) *Arabidopsis thaliana* DNA methylation mutants. *Science* **260**, 1926-1928.

Summary

In this thesis, using HR-promoted GT, identical mutations for the rice *ROS1a* gene encoding a putative cytosine DNA demethylase and for the rice *DDM1a/DDM1b* genes encoding SWI2/SNF2 chromatin remodeling factors were reproducibly generated. Obtained knock-in/knock-out mutants have shown that proper cytosine demethylation is indispensable in order to facilitate normal developmental processes in rice life cycle, and also that proper cytosine methylation is necessary for protection of rice genome from undesired events, such as potentially harmful reactivation of TEs.

Chapter 1

Genes that promote DNA methylation and demethylation in plants have been characterized mainly in Arabidopsis. Arabidopsis DNA demethylation is mediated by bifunctional DNA enzymes with glycosylase activity that removes 5-methylcytosine and lyase activity that nicks the double-stranded DNA at the abasic site. Homologous recombination–promoted knock-in targeting of the *ROS1a* gene, the longest of six putative DNA demethylase genes in the rice genome, could repeatedly disrupt *ROS1a* in primary (T_0) transgenic plants by fusing its endogenous promoter with the *GUS* reporter gene in the heterozygous

condition. These T_0 plants exhibited no overt morphological phenotypes during their vegetative phase, and GUS staining showed *ROS1a* expression in pollen, unfertilized ovules, and meristematic cells. Interestingly, neither the maternal nor the paternal knock-in null allele, *ros1a-GUS1*, was detected in the progeny; such an intransmittable null mutation would be difficult to isolate by conventional mutagenesis techniques that usually identify and isolate mutants in the progeny population. Even in the presence of the wild-type paternal *ROS1a* allele, the maternal *ros1a-GUS1* allele caused failure of early stage endosperm development resulting in incomplete embryo development, with embryogenesis producing irregular but viable embryos that could not complete seed dormancy, implying nonequivalent maternal and paternal contribution of ROS1a in endosperm development. The paternal *ros1a-GUS1* allele could not be transmitted to progeny, presumably because of a male gametophytic defect(s) prior to fertilization. Thus, *ROS1a* must be indispensable in both male and female gametophytes, in which DNA demethylation must play important roles.

Chapter 2

DNA methylation plays important roles in regulating gene expression. Using homologous recombination-promoted knock-out targeting, an identical mutation for each of the rice *DDM1a* and *DDM1b* genes encoding putative SWI2/SNF2-like

chromatin remodeling ATPases were reproducibly generated. No effect of *OsDDM1a* disruption on cytosine methylation was observed at any repetitive sequences investigated. In contrast, *OsDDM1b* disruption caused decreased cytosine methylation in repetitive sequences. Subsequently generated *OsDDM1a* and *OsDDM1b* double disruptant mutants exhibited drastic reduction of cytosine methylation indicating the different degree of contribution and the functional redundancy in certain extent between OsDDM1a and OsDDM1b. Double disruption derepressed the transcription of several transposable elements suggesting an essential role for OsDDM1s on epigenetic regulation of repetitive sequences that comprise a large component of rice genome and have contributed to large-scale genome insertions, rearrangements and allelic diversification during its evolution.

Acknowledgement

I am deeply grateful to my advisors Professor Toshiaki Mitsui and Associate Professor Kimiko Itoh, and my external reviewer Professor Shigeru Iida for their constant guidance, encouragement, and scientific suggestions throughout this study. I am very grateful to other members of my committee, Professors Takuji Ohyama and Keiichi Okazaki for their insightful comments and kind encouragement. I would like to acknowledge Professors Martin G. Marinus and Robert L. Fischer for providing *E. coli* strains and plasmids, and to Yoko Mizuta and Saori Miyazaki for their valuable suggestions on the *in vitro* pollen germination assay. I would like to thank all the members of Molecular Genetics lab in National Institute for Basic Biology (NIBB) and the NIBB Center for Analytical Instruments for their technical support and valuable discussions. In particular, I would like to thank Katsushi Yamaguchi, Yasuyo Johzuka-Hisatomi, Sachiko Fukada-Tanaka, and Rie Terada for their technical assistance, valuable suggestions, and invaluable help. I would also like to thank Kazuo Tsugane for all the help on microarray experiments and Takaki Yamauchi for his tremendous assistance on the analysis of microarray experiment and the fruitful discussions. Last but not least, I would like to thank my family who has always been supportive and encouraging in times of difficulties.