Support Information

Mutations in the DNA demethylase *OsROS1* result in a thickened aleurone and improved nutritional value in rice grains.

Jinxin Liu, Xiaoba Wu, Xuefeng Yao, Ronald Yu, Philip J. Larkin, Chun-Ming Liu

SI Materials and Methods

Plant materials and growth conditions

Rice (*Oryza sativa* L. ssp. *Geng*, cultivar Zhonghua 11, ZH11) plants were cultivated in an experimental field at the Institute of Botany, Chinese Academy of Sciences, Beijing, or growth chambers with 60% humidity and daily light of 12 h, 30 ± 2℃ during the day, and 22 ± 2℃ during the night. These chambers were equipped with Philips GreenPower LED Toplighting lamps (High Output System, 410 μmol/s). The *ta2-1* mutant was identified by screening M3 seeds produced from an M2 population of plants mutagenized using ethyl methanesulfonate (EMS) and was backcrossed to ZH11 for at least five generations.

Half-seed assays and histological analyses in hand-cut sections

For Evans Blue staining-based half-seed assay, dehusked mature grains were sectioned transversally into two halves using a razor blade, immersed in 0.1% (w/v) Evans Blue solution for 10 min, and washed three times with distilled water. For Sudan IV/Lugol staining, dehusked mature grains sectioned in the same manner were stained first in 0.5% (w/v) Sudan IV for 24 h at room temperature, and then in Lugol's iodine (Sigma) for 1 min before being washed three times with distilled water. All samples were observed and photographed under a dissecting microscope (SMZ800, Nikon).

Histological analysis of semi-thin sections

Rice grains collected at different stages of development were sectioned transversely into 2-mm slices using a razor blade, and fixed immediately in a modified FAA solution (50% ethanol, 5% formaldehyde, 6% acetic acid, 5% glycerol) for 24 h, followed by dehydration and embedding in L. R. White resin (The London Resin Company, UK) as described (1). Semi-thin sections (1.25 μm) were cut using a glass knife in a Leica ultra-microtome (EM UC7, Leica Microsystems), stained with periodic acid-Schiff (PAS) reagent, and counter-stained with 0.1% (w/v) Coomassie Brilliant Blue (CBB) for 15 min. These sections were then washed three times with double distilled water before mounting with a coverslip. Photographs were taken using a Nikon microscope (Eclipse 80i).

Measurement of nutritional contents

Total proteins were measured according to the NY/T3-1982 standard of determination of crude protein content in cereals and legumes (semimicro-Kjeldahl method). Total lipids were measured according to the AOAC official method 983.23. Iron, zinc, and calcium contents were measured by inductively coupled plasma optical emission spectrometry (ICP-OES) as described (2). Dietary fiber was measured according to the AOAC official method 985.29.

Antioxidants were measured following a previously reported method (3). Total starch content was analyzed using the Total Starch Measurement Kit (Megazyme International Ireland Ltd.). Amylose content was measured using the Amylose/Amylopectin Assay Kit (Megazyme International Ireland Ltd.). Vitamins A, B1, B2, B3, B6, and E were measured by the Cereal Quality Supervision and Monitoring Center, Ministry of Agriculture, China.

Analyses of agronomic traits

Major agronomic traits, including plant height, panicle length, tiller number, and the number of primary branches in the main panicle and number of grains in the main panicle of ZH11 and *ta2-1* and *ta2* TILLING alleles, were measured in plants grown in the abovementioned experimental field. Grain length, width, and thickness, and 1,000-grain weight were measured in mature grains. Grain yield per plant was measured in 20–30 plants grown in three randomized blocks in the same field. The actual yields of ZH11 and *ta2-1* were measured in 4 m by 4 m plots.

Map-based cloning

To map the *TA2* gene, the *ta2-1* mutant was crossed with Nanjing 6, a *Xian* rice variety, which allowed the *TA2* locus to be rough mapped between the InDel markers L19 and L32 using 50 homozygous *ta2-1* individuals identified in the F2 generation. Using another 2,128 wild-type plants identified in over 7,000 F2 plants, *TA2* was fine-mapped to a 46-kb genomic region between markers S7 and S18.

Molecular constructions and plant transformation

For genetic complementation, a 16.2-kb genomic fragment spanning the entire coding region and the 4.7-kb 5′ upstream regions of *OsROS1* was amplified from ZH11 plants and inserted into *pCAMBIA1300* to generate the *pCAMBIA1300-OsROS1* complementation construct using the EasyGeno Assembly Cloning Kit (Tiangen). The complementation construct was transformed into homozygous *ta2-1*. To construct *pUbi*::*OsROS1* and *pUbi*::*mOsROS1*, full-length *OsROS1* and *mOsROS1* cDNAs were amplified from ZH11 and *ta2-1*, respectively, cloned into the binary vector *pCAMBIA1301-Ubi* using the EasyGeno Assembly Cloning Kit (Tiangen), and confirmed by sequencing before being transformed into ZH11. All transgenic plants were generated by *Agrobacterium*-mediated transformation as described (4).

RNA *in situ* hybridization

A 507-bp gene-specific region of the 3′ UTR of *OsROS1* was amplified and sub-cloned into a *pEASY-Blunt* vector (Transgen) to prepare the sense and antisense probes. RNA *in situ* hybridization was performed as reported (5).

RNA extraction, cDNA synthesis, and qRT-PCR

Total RNA extraction and cDNA synthesis were performed in various tissues using Trizol reagent (Invitrogen) and a FastQuant RT Kit (Tiangen), respectively, according to the manufacturers' instructions. RT-qPCR was performed using 10 ng cDNA, 0.2 μM primer, and SuperReal PreMix Plus (Tiangen) in the CFX96 real-time system (Bio-Rad) according to the manufacturers' instructions. The rice ubiquitin transcript was used as the internal control.

Three independent replicates were analyzed for each sample. Primer sequences are provided (*SI Appendix* Table S4).

Phylogenetic analysis

Homologs of OsROS1 were identified from GenBank using a BLASTP search (http://www.ncbi.nlm.nih.gov). The phylogenetic tree was constructed using MEGA 7.0 (http://www.megasoftware.net).

Identification of new *ta2* alleles by TILLING

New *ta2* alleles were identified through an in-house TILLING platform (www.croptilling.org) by screening the region from 6,444,300 to 6,449,200 at chromosome 1. New *ta2* alleles identified were backcrossed to ZH11 twice before phenotypic analyses were performed.

WGBS library construction and data analysis

For whole-genome bisulfite sequencing, genomic DNA was isolated from endosperms and embryos using a QIAamp DNA Mini Kit (Qiagen). The volumes of the combined gDNA sample and unmethylated lambda DNA control were adjusted to a total volume of 80 μL using 1x TE buffer, and fragmented to 300-bp. Subsequently, fragmented DNA was end repaired and ligated to methylated adapters, and a gDNA library was constructed according to the manufacturer's instructions (Illumina). Bisulfite conversion was performed using an EZ DNA Methylation-Gold Kit. The bisulfiate-converted library was amplified using KAPA HiFi HotStart DNA Polymerase (Kapa Biosystems). Sequencing on an Illumina Hiseq X Ten sequencer, as paired-end 150-bp reads, was performed at Annoroad Gene Technology Company.

For WGBS data analysis, low-quality reads were filtered. Clean reads were aligned to a reference genome (Os-Nipponbare-Reference-IRGSP-1.0) using Bismark (6). Total reads, aligned reads, and genomic coverage are listed (*SI Appendix*, Table S5). Methylation levels were calculated as mC/(mC+C). DMRs were identified by comparing the methylation levels of CG, CHG, and CHH with those of ZH11 (7). The reference genome was divided into 100-bp bins for comparing the number of methylated and unmethylated Cs with at least three Cs covered. Absolute methylation differences of each bin had to be at least 0.1 for all three contexts (Fisher's exact test p-value < 0.00005). For patterns of DNA methylation in gene bodies, TEs and their flanking regions, methylation levels within a 100-bp interval of the genome were averaged. Two-kilobase regions upstream and downstream of the genes and TEs were divided into 20 equal bins, and 4-kb into the annotated region of genes and TEs were divided into 40 equal bins.

References

1. Wu X, Liu J, Li D, Liu CM (2016) Rice caryopsis development I: Dynamic changes in different cell layers. *J Integr Plant Biol* 58: 722-785.
2. Zarcinas BA, Cartwright B, Spouncer LR (1987) Nitric acid digestion and multi−element analysis of plant material by inductively coupled plasma spectrometry. *Commun Soil Sci Plan* 18: 131-146.

3. Huang D, Ou B, Hampsch-Woodill M, Flanagan JA, Prior RL (2002) High-throughput assay of oxygen radical absorbance capacity (ORAC) using a multichannel liquid handling system coupled with a microplate fluorescence reader in 96-well format. *J Agric Food Chem* 50: 4437-4444.
4. Nishimura A, Aichi I, Matsuoka M (2007) A protocol for *Agrobacterium*-mediated transformation in rice. *Nat Protoc* 6: 2796-2802.
5. Brewer PB, Heisler MG, Hejátko J, Friml J, Benková E (2006) *In situ* hybridization for mRNA detection in *Arabidopsis* tissue sections. *Nat Protoc* 3: 1462-1467.
6. Krueger F, Andrews SR (2011) Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* 27: 1571-1572.
7. Wu H, et al. (2015) Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Res* 43: e141.

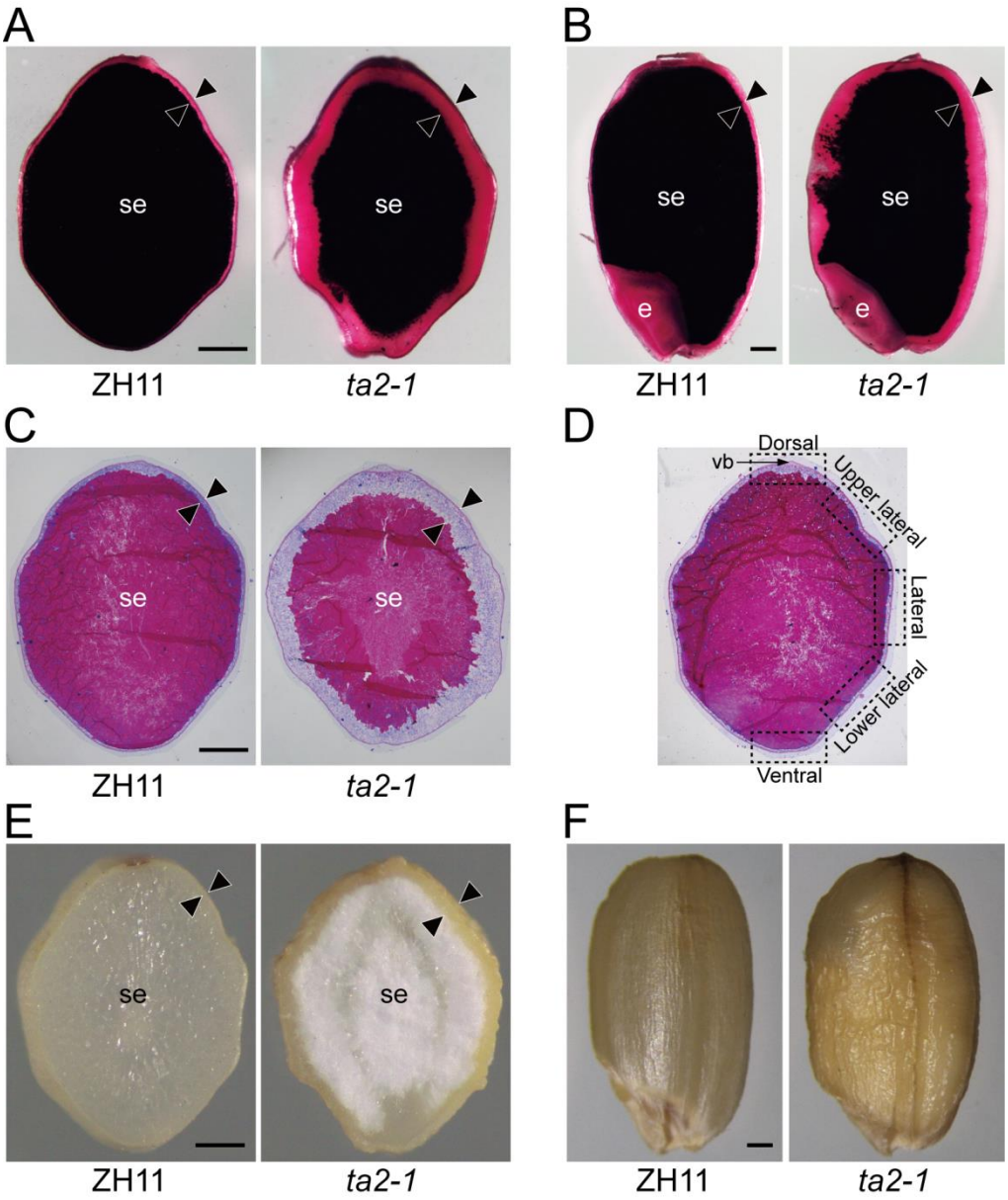

Fig. S1. The *ta2-1* mature grains exhibited thickened aleurones. (*A* and *B*) Transversely (*A*) and longitudinally (*B*) sectioned ZH11 and *ta2-1* dehusked mature grains, counter-stained with Sudan Red IV (red) for lipids, and Lugo iodine (black) for starch. Note that lipids accumulated mainly in the aleurone and embryo, while starch accumulated mainly in the endosperm. (*C*) Semi-thin sections of ZH11 and *ta2-1* dehusked mature grains, stained by periodic acid-Schiff (PAS) reagent and Coomassie Brilliant Blue (CCB). (*D*) A transversely sectioned dehusked mature grain showing the dorsal, upper lateral, lateral, lower lateral, and ventral positions of the endosperm described in this study. (*E* and *F*) Transversely sectioned (*E*) and non-sectioned (*F*) mature grains showing the opaque endosperm in *ta2-1*, as compared to the semi-transparent endosperm in ZH11. (Scale bar, 0.5 mm.) Arrowheads indicate the aleurone. se, starchy endosperm; e, embryo; vb, vascular bundle.

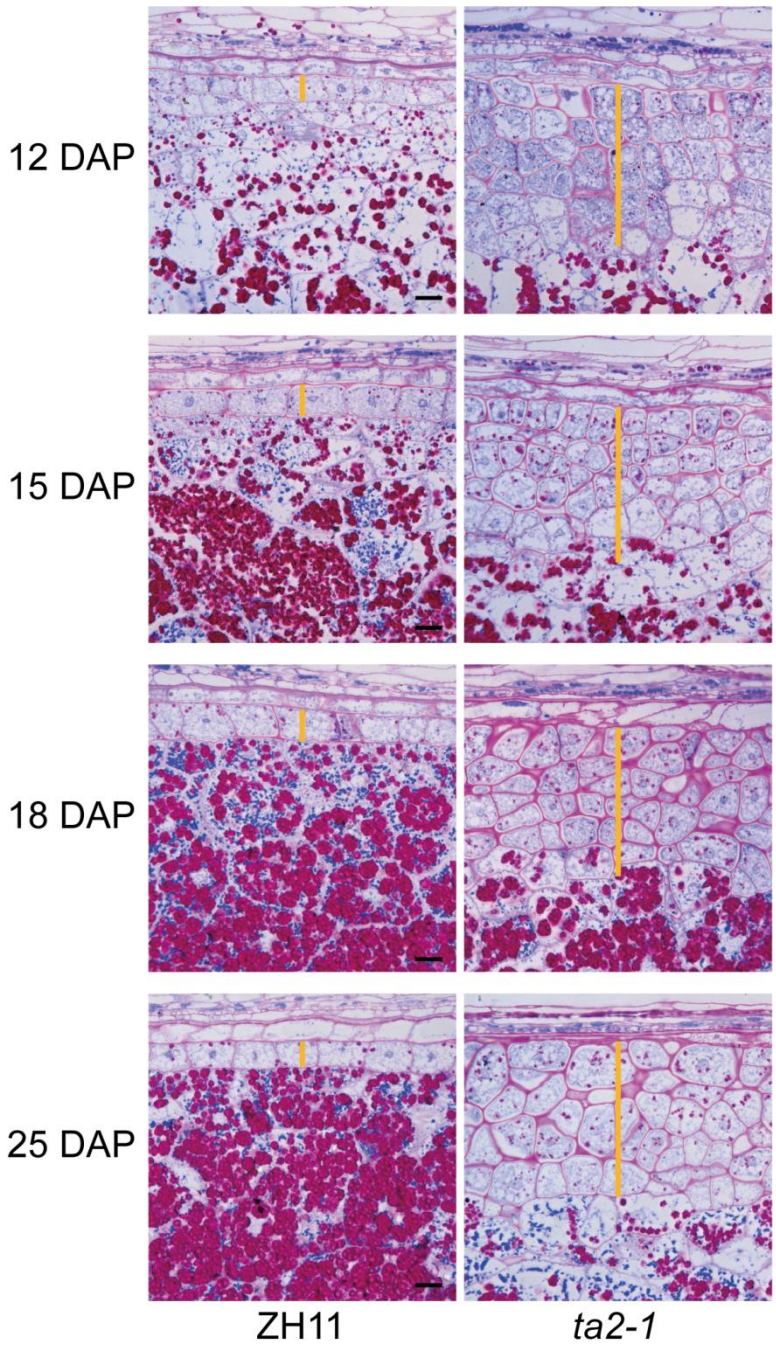

Fig. S2. Aleurone at the lateral position of ZH11 and *ta2-1*. Photos were taken from semi-thin sectioned ZH11 and *ta2-1* endosperms, stained with PAS and CBB, at 12, 15, 18, 25 DAP. Orange lines indicate the aleurone. (Scale bar, 20 μm.)

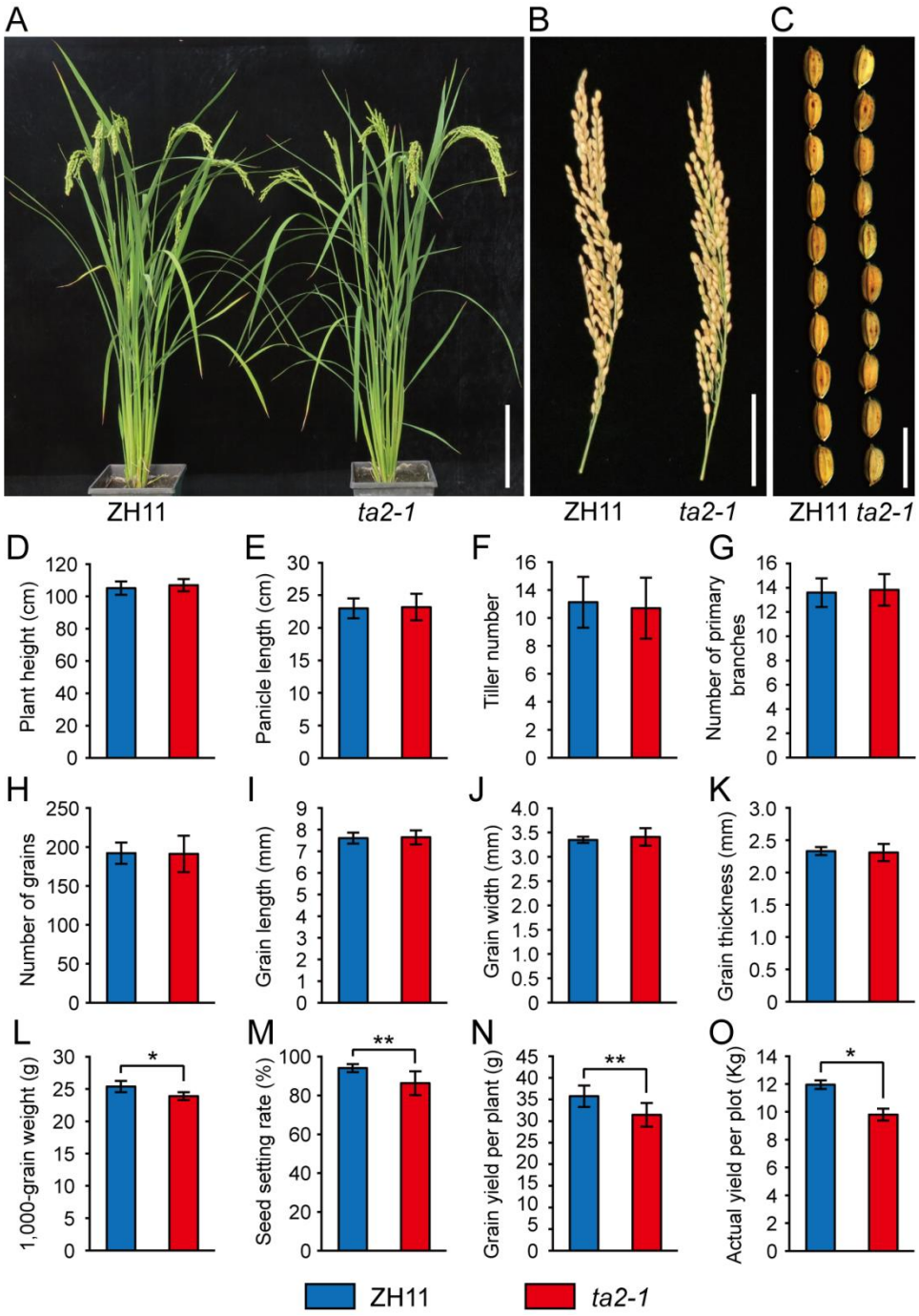

Fig. S3. Yield-related traits of ZH11 and *ta2-1* plants. (*A*) ZH11 and *ta2-1* plants at anthesis. (Scale bar, 20 cm.) (*B*) Mature ZH11 and *ta2-1* panicles. (Scale bar, 5 cm.) (*C*) Length of ZH11 and *ta2-1* grains. (Scale bar, 1 cm.) (*D*) Plant height. (*E*) Panicle length. (*F*) Tiller number per plant. (*G*) Number of primary branches in the main panicle. (*H*) Number of grains in the main panicle. (*I*) Grain length. (*J*) Grain width. (*K*) Grain thickness. (*L*) 1,000-grain weight. (*M*) Seed setting rate. (*N*) Grain yield per plant. (*O*) Actual yield per plot. Data are means ± SD (*D-K*, n = 40; *L-N*, n = 30; *O*, n = 3). * and ** denote significant differences from ZH11 at the 0.05 and 0.01 levels, respectively, according to Student's *t*-tests.

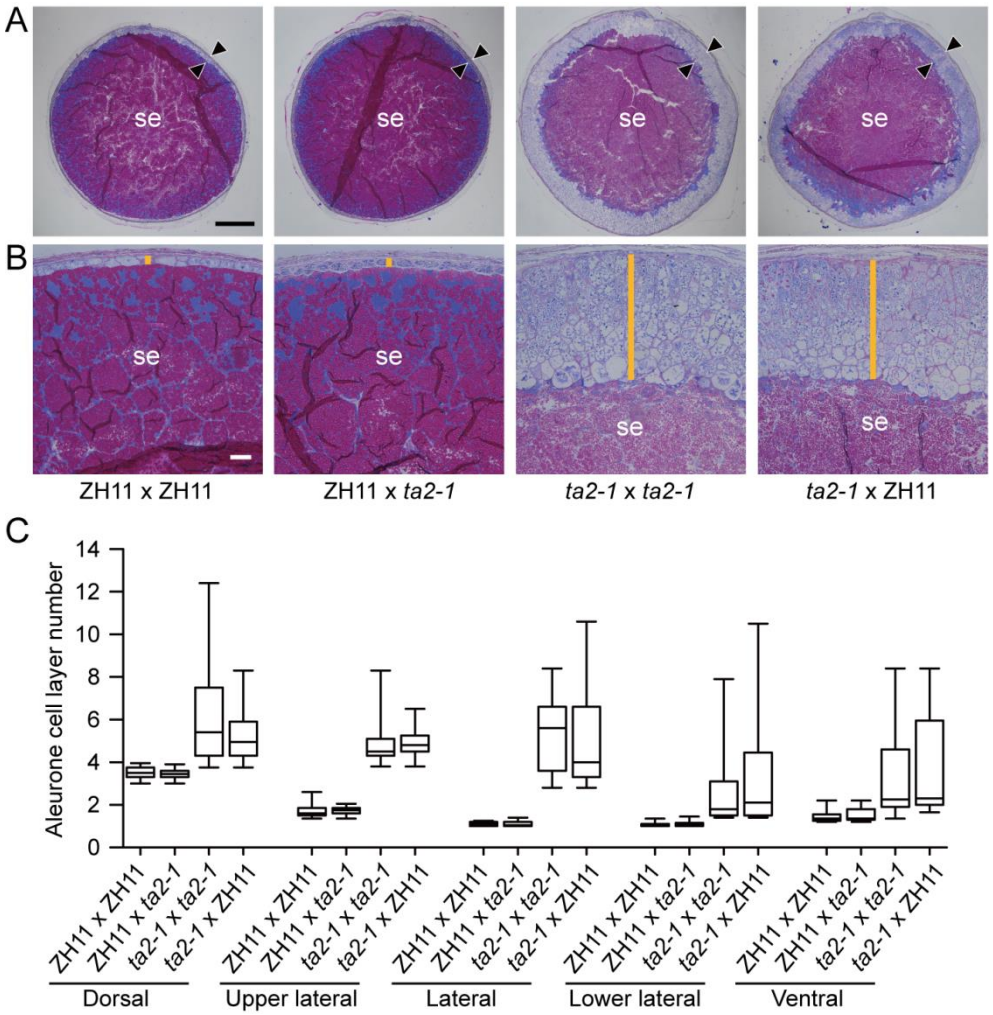

Fig. S4. Aleurone phenotypes after reciprocal crosses between ZH11 and *ta2-1*. (*A*) Semi-thin transversally sectioned dehusked grains after reciprocal crosses between ZH11 and *ta2-1*. Arrowheads indicate the aleurone. (Scale bar, 0.5 mm.) (*B*) Morphologies of the aleurone at the lateral position in different crosses, with the female parent written first. Orange lines indicate the aleurone. se, starchy endosperm. (Scale bar, 50 μm.) (*C*) Box plots of the average numbers of aleurone cell layers in various parts of the endosperm in mature grains of different crosses (n = 10).

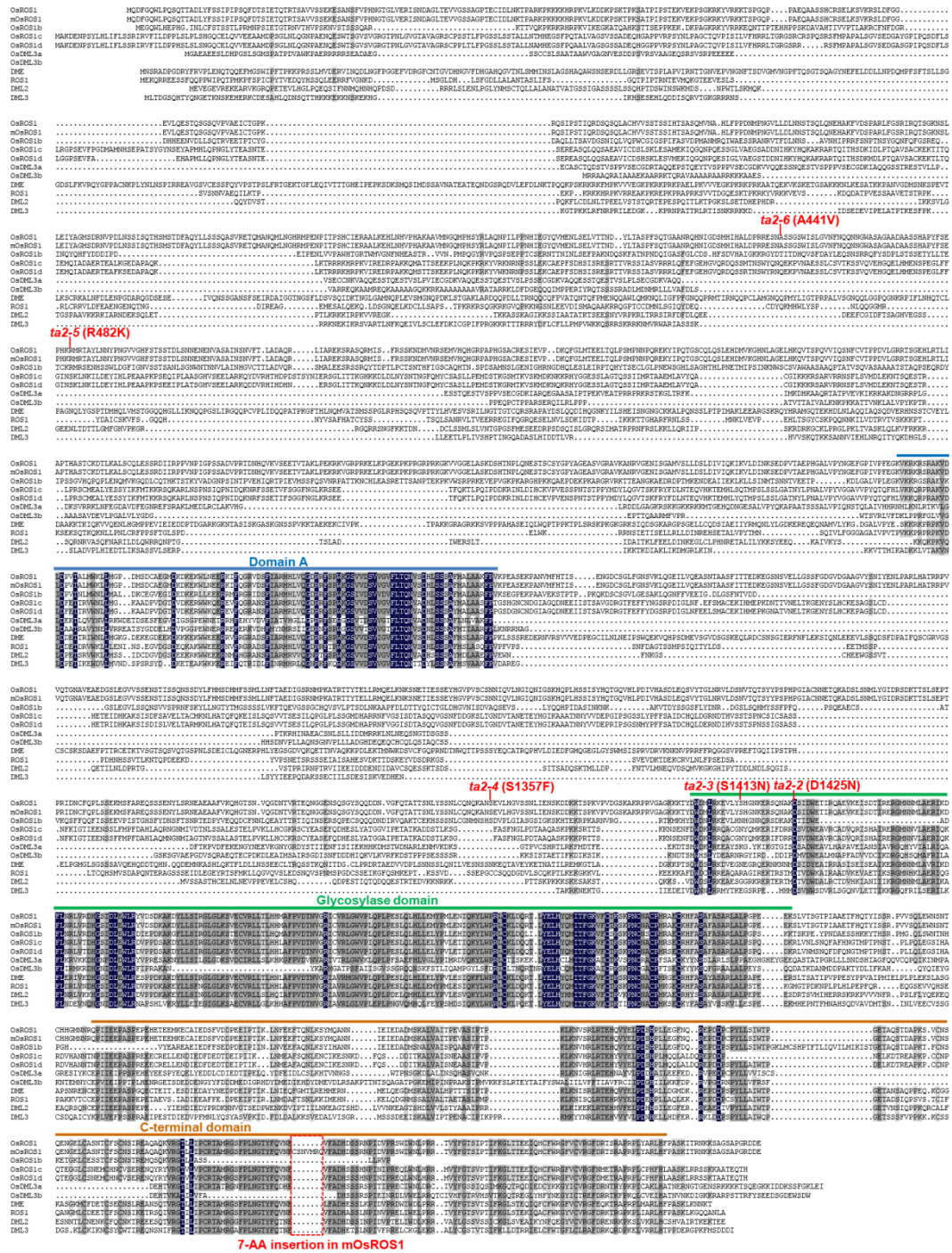

Fig. S5. Alignment of full-length OsROS1 with its homologs in rice and Arabidopsis. The alignment was generated by DNAMAN using ClustalX. Rice: OsROS1 (Os01g0218032); OsROS1b (Os02g0494700); OsROS1c (Os05g0445900); OsROS1d (Os05g0446600); OsDML3a (Os02g0496500); OsDML3b (Os04g0357850). Arabidopsis: DME (AT5g04560); ROS1 (AT2g36490); DML2 (AT3g10010); DML3 (AT4g34060). Positions of three conserved domains, the domain A (blue line), the glycosylase domain (green line), and the C-terminal domain (brown line), are indicated above the alignment. The red dashed box indicates the 7-AA insertion in mOsROS1 of *ta2-1*, and mutations in five TILLING alleles are marked in red.

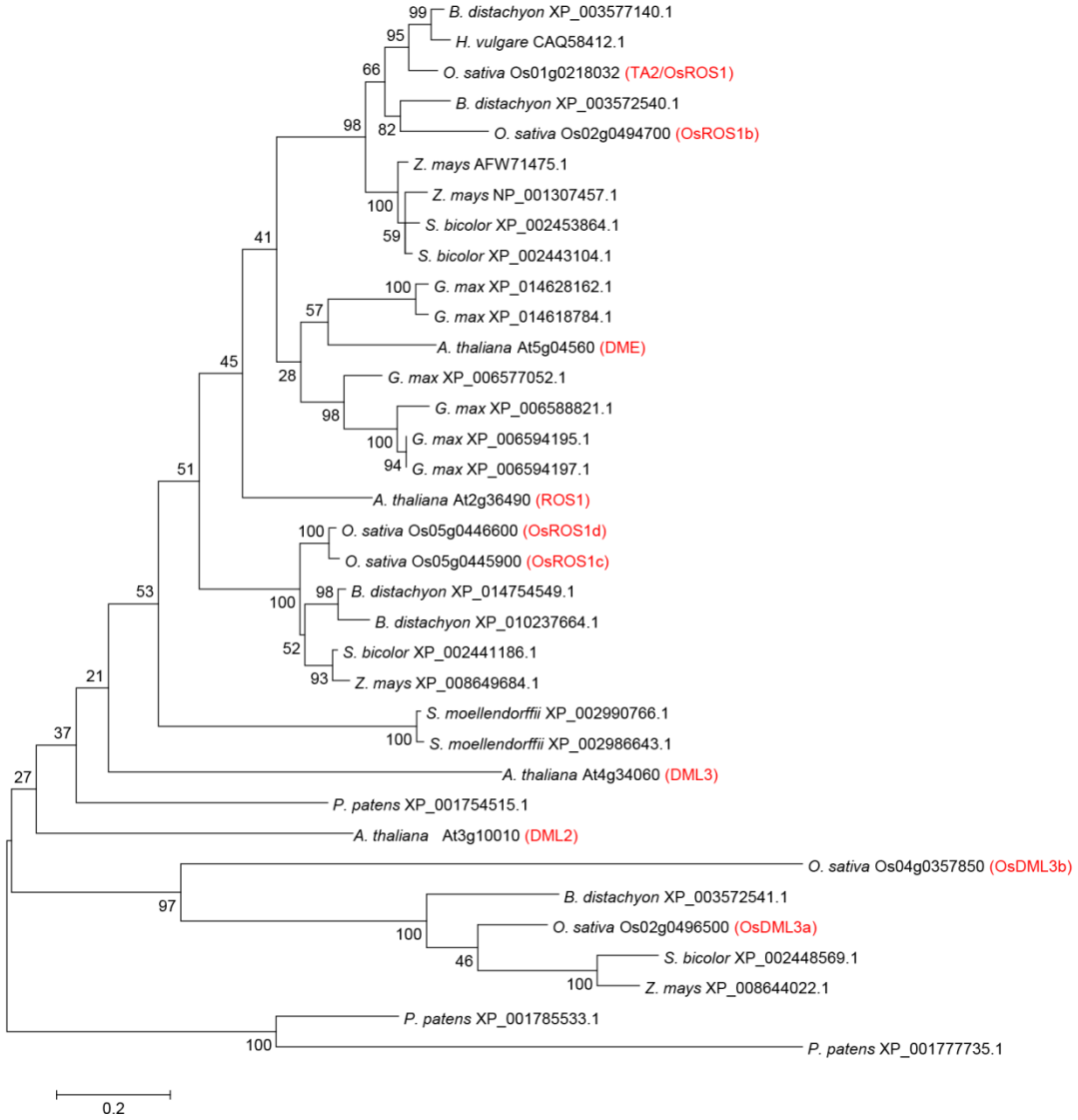

Fig. S6. Phylogenetic analysis of OsROS1 and its homologs from monocot and dicot plants. Multiple sequence alignment was performed using Clustalw2. Maximum-likelihood phylogenetic tree was constructed using MEGA7. Data was showed based on a cutoff *E* value of < 0.1. Numbers represent support based on 100 bootstrap replicates. Branch length represents substitutions per site.

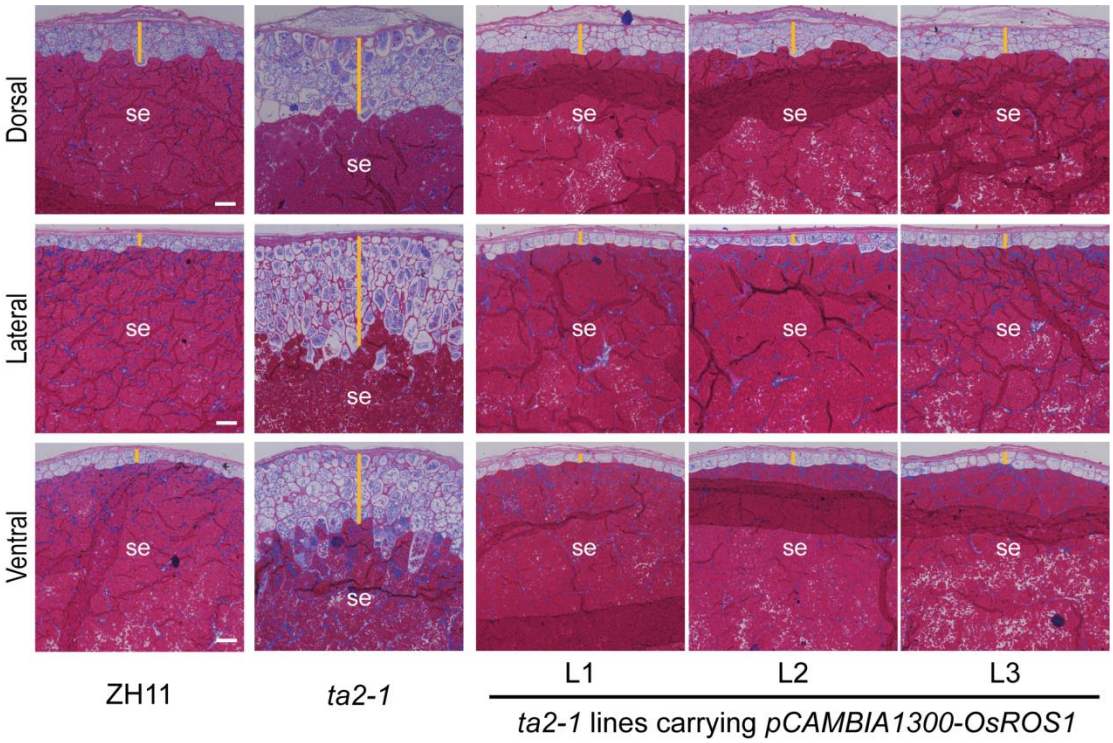

Fig. S7. Complementation analysis performed in *ta2-1*. Semi-thin sectioned dehusked mature grains of ZH11, *ta2-1*, and three *ta2-1* transgenic lines (L1, L2, and L3) carrying *pCAMBIA1300-OsROS1*, stained with PAS and CBB, to show complete complementation of the *ta* phenotype. Photographs were taken at the dorsal, lateral, and ventral positions of the endosperm. Orange lines indicate the aleurone. se, starchy endosperm. (Scale bar, 50 μm.)

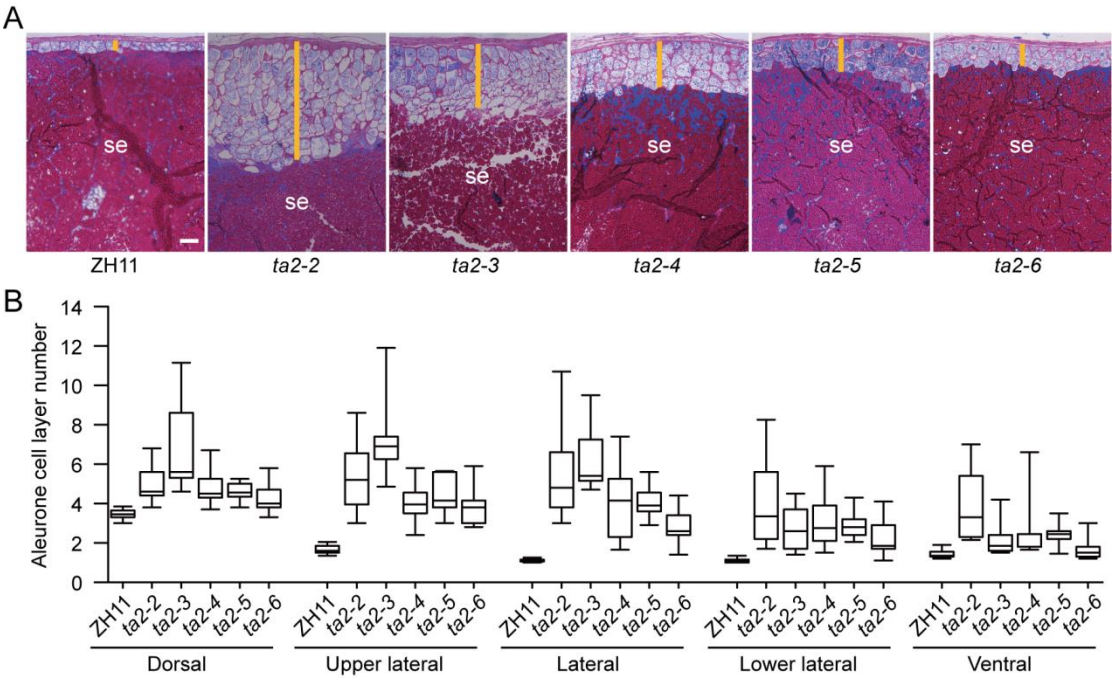

Fig. S8. Morphologies of the aleurone in ZH11 and five *ta2* TILLING alleles. (*A*) Morphologies of the aleurone in ZH11 and *ta2-2* to *ta2-6*, at the lateral position. Orange lines indicate the aleurone. se, starchy endosperm. (Scale bar, 50 μm.) (*B*) Average number of aleurone cell layers in ZH11, and *ta2-2* to *ta2-6*, in different parts of the endosperm, shown in box plots (n = 10).

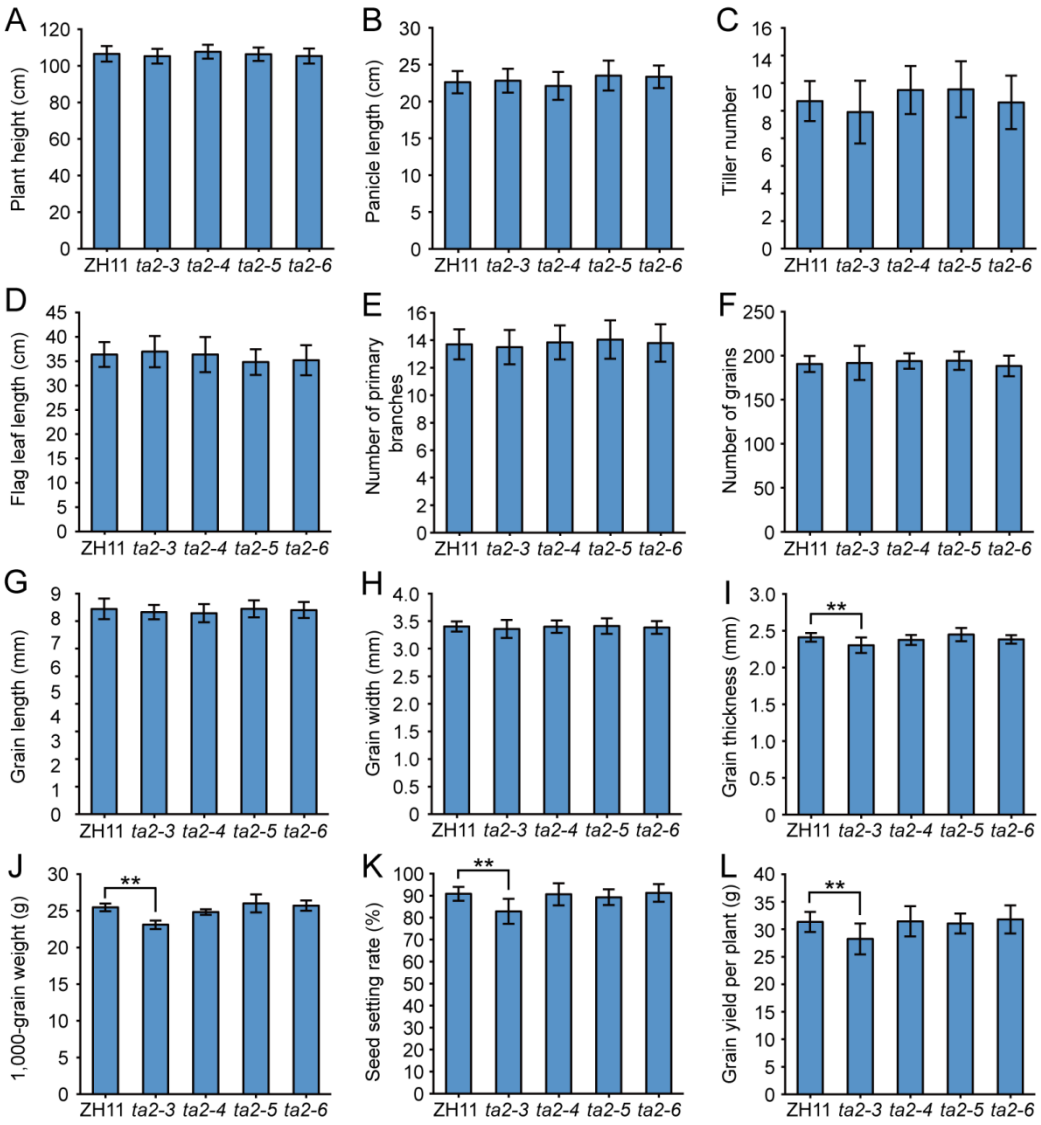

Fig. S9. Yield-related traits of four *ta2* TILLING alleles. (*A*) Plant height. (*B*) Panicle length. (*C*) Tiller number per plant. (*D*) Flag leaf length. (*E*) Number of primary branches in the main panicle. (*F*) Number of grains in the main panicle. (*G*) Grain length. (*H*) Grain width. (*I*) Grain thickness. (*J*) 1,000-grain weight. (*K*) Seed setting rate. (*L*) Grain yield per plant. Data are means ± SD (n = 20). ** denote significant differences from ZH11 at the 0.01 levels according to Student's *t*-tests.

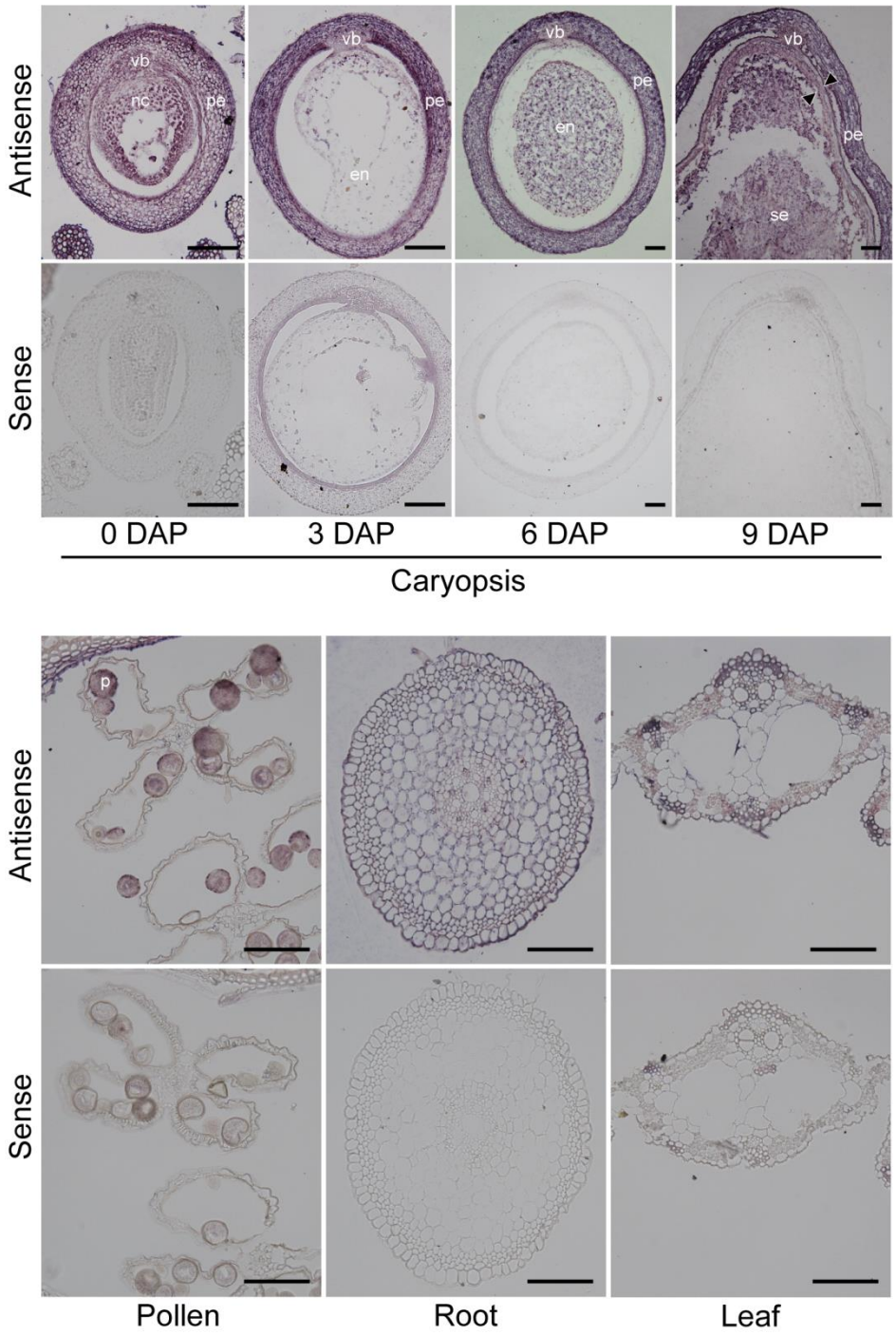

Fig. S10. Analyses of *OsROS1* expression. *In situ* hybridization was performed to show the expression of *OsROS1* in caryopses at 0, 3, 6, and 9 DAP, and pollen, root and leaf. (Scale bars, 100 μm.) Arrowheads indicate the aleurone. vb, vascular bundle; nc, nucellar cell; pe, pericarp; en, endosperm; p, pollen; se, starchy endosperm.

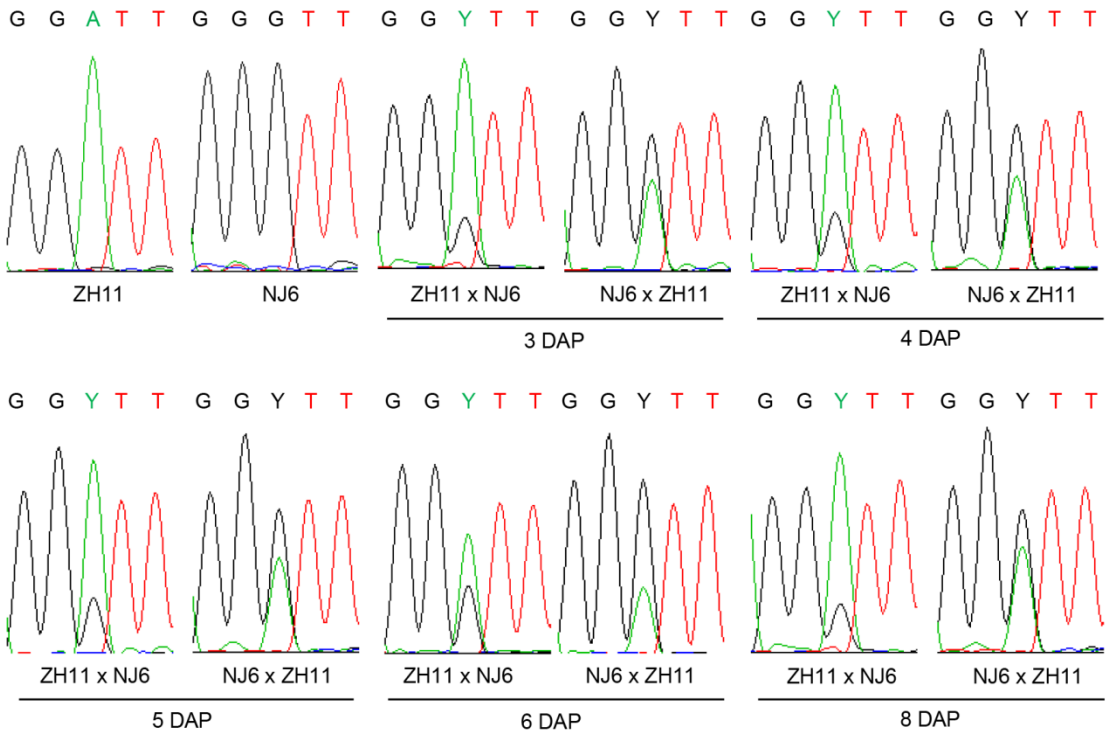

Fig. S11. Expression of paternal and maternal *OsROS1* alleles. Direct sequencing of RT-PCR products was used to examine the expressions of paternal and maternal *OsROS1* alleles in cDNA prepared from endosperms collected at 3, 4, 5, 6, and 8 DAP in reciprocal crosses between ZH11 and NJ6. A SNP, A or G, located in the amplicon was used to discriminate the maternal and paternal *OsROS1* alleles. Y represents the presence of both A and G.

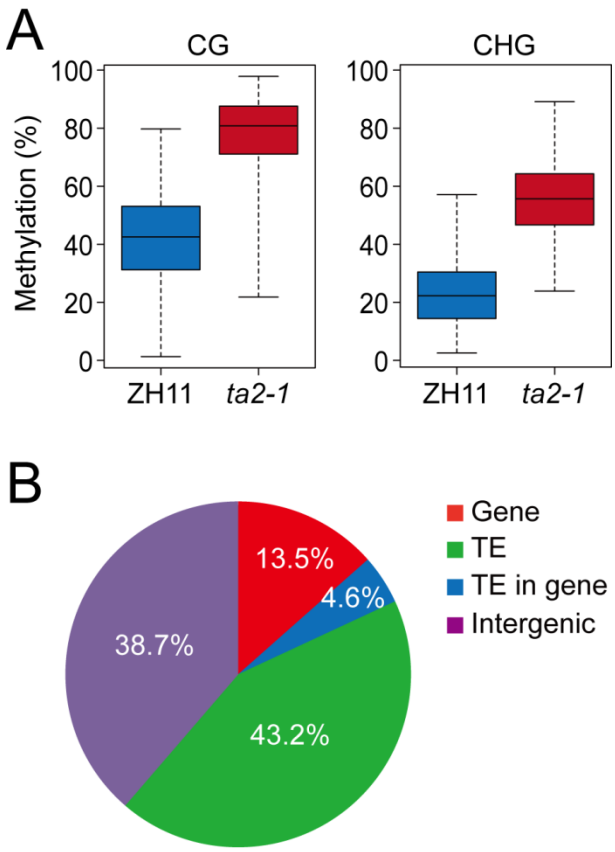

Fig. S12. Analysis of hyper-DMRs in *ta2-1* endosperm. (*A*) Box plots displaying the average methylation levels of CG and CHG hyper-DMRs in ZH11 and *ta2-1* endosperms at 9 DAP. (*B*) Composition of hyper-DMRs in *ta2-1* endosperms in gene body, intergenic, TE, and TE and gene overlapping regions.

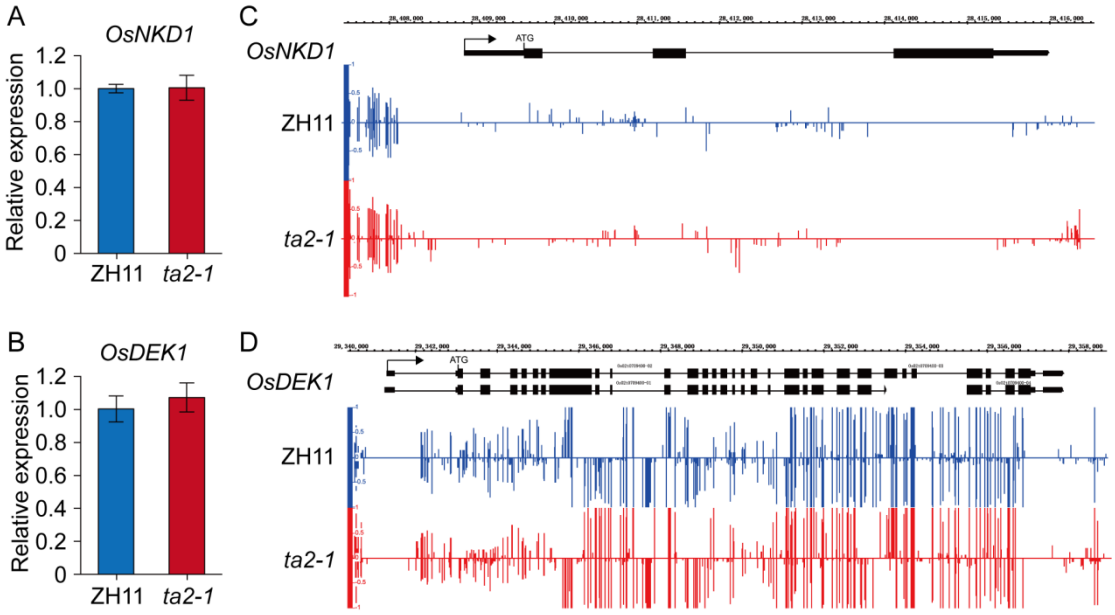

Fig. S13. The DNA methylation and expression levels of *OsNKD1* and *OsDEK1*. (*A* and *B*) No significant differences were observed in *OsNKD1* (*A*) and *OsDEK1* (*B*) expression levels between ZH11 and *ta2-1* endosperms at 9 DAP, based on qRT-PCR analyses. Data are means ± SD (n = 3). (*C* and *D*) Snapshots in the Integrated Genome Browser, showing DNA methylation levels of the *OsNKD1* (*C*) and *OsDEK1* (*D*) in ZH11 and *ta2-1* endosperms at 9 DAP.

Table. S1. Percentage of amino acid sequence identities of OsROS1 with its homologs in Arabidopsis and rice.

Protein	**Plant species**	**Identity**
OsROS1c	*Oryza sativa*	68%
ROS1	*Arabidopsis thaliana*	67%
OsROS1b	*Oryza sativa*	66%
DME	*Arabidopsis thaliana*	63%
DML2	*Arabidopsis thaliana*	57%
DML3	*Arabidopsis thaliana*	45%

Amino acid sequence identities were calculated based on alignments of the glycosylase and C-terminal domains.

Table. S2. List of 23 TILLING alleles of *OsROS1*.

Allele name	**Mutation in protein**	**Aleurone phenotype**
ta2-2	D1425N	Thick aleurone
ta2-3	S1413N	Thick aleurone
ta2-4	S1357F	Thick aleurone
ta2-5	R482K	Thick aleurone
ta2-6	A441V	Thick aleurone
ta2-7	D3V	Wild type
ta2-8	A78V	Wild type
ta2-9	E123K	Wild type
ta2-10	S214F	Wild type
ta2-11	T221I	Wild type
ta2-12	R530K	Wild type
ta2-13	P843C	Wild type
ta2-14	P883S	Wild type
ta2-15	P1225L	Wild type
ta2-16	S1272N	Wild type
ta2-17	R1390N	Wild type
ta2-18	synonymous	Wild type
ta2-19	synonymous	Wild type
ta2-20	synonymous	Wild type
ta2-21	synonymous	Wild type
ta2-22	synonymous	Wild type
ta2-23	synonymous	Wild type
ta2-24	synonymous	Wild type

Table S3. Genetic analyses of four representative *ta2* TILLING alleles.

Cross combination	Endosperm phenotypes		Grains with *ta* phenotype (%)	*P* for 1:1	*P* for 3:1
	Wild type	*ta*			
ta2-3 x ZH11	0	172	100	N.A.	N.A.
ZH11x *ta2-3*	70	2	2.8	N.A.	N.A.
TA2/*ta2-3* x ZH11	65	67	50.8	0.8618 (N.S.)	N.A.
ZH11x *TA2*/*ta2-3*	156	0	0	N.A.	N.A.
ta2-4 x ZH11	79	68	46.3	0.3643 (N.S.)	N.A.
ZH11 x *ta2-4*	69	0	0	N.A.	N.A.
TA2/*ta2-4* x ZH11	59	15	20.1	N.A.	0.3474 (N.S.)
ZH11 x *TA2*/*ta2-4*	87	0	0	N.A.	N.A.
ta2-5 x ZH11	219	76	25.8	N.A.	0.0576 (N.S.)
ZH11 x *ta2-5*	113	0	0	N.A.	N.A.
TA2/*ta2-5* x ZH11	150	30	16.7	N.A.	0.0098 (S.)
ZH11 x *TA2*/*ta2-5*	208	0	0	N.A.	N.A.
ta2-6 x ZH11	116	25	17.7	N.A.	0.0462 (S.)
ZH11 x *ta2-6*	173	0	0	N.A.	N.A.
TA2/*ta2-6* x ZH11	58	8	12.1	N.A.	0.0157 (S.)
ZH11 x *TA2*/*ta2-6*	224	0	0	N.A.	N.A.

N.A., not applicable; N.S., not significant; S., significant.

Table S4. Primers used in this study.

Primer	Sequence (5' to 3')	Purpose
ta2 gDNA mut-F	TAGATACCATGCCGAACAGC	Genotyping
ta2 gDNA mut-R	GTGTACTTGACATTTCAAATCACAG	
ta2 cDNA mut-F	CGAGGGACACTGCTGATA	Genotyping
ta2 cDNA mut-R	CCAAAGTAAACAGTTCTCCTAG	
OsROS1-gDNA-F	AGGGCATTGTAGACATTAGCATTTC	Genetic complementation
OsROS1-gDNA-R	CCTAACAACATGAATGACCATTTTG	
OsROS1 imprinting F	TGAAACAGCTCAATCAACTGATGCACCTA	Gene imprinting testing
OsROS1 imprinting R	CTATTCATCATCTCTTCCTGGAGCAGAACC	
OsROS1 cDNA-F	ATGACATGATTACGAATTCGGTACCATGCAGGATTTTGGACAATGGCTGC	*pUbi:OsROS1* and *pUbi:mOsROS1* construction
OsROS1 cDNA-R	AATGTTTGAACGCCTGCAGGTCGACCTATTCATCATCTCTTCCTGGAGCAGAACC	
OsROS1-qPCR-F	GTTCGAGGGACACTGCTGATAC	qPCR for m*OsROS1*
mOsROS1-qPCR-R	TGCCGCATAACATTTGAACAC	
OsROS1-qPCR-F	GTTCGAGGGACACTGCTGATAC	qPCR for *OsROS1*
OsROS1-qPCR-R	TTGAGTCATGATCAGCAAATACC	
RISBZ1-qPCR-F	CAAGGGAGCCATCACCATC	qPCR for *RISBZ1*
RISBZ1-qPCR-R	CCTTTCTGCTTCTTGAGCGTCTA	
RPBF -qPCR-F	ACTACGCGCCTCTCATCACC	qPCR for *RPBF*
RPBF -qPCR-R	TCACTCCACCACCACCTCCT	
OsNKD1-qPCR-F	CTCCTTGGAGGCCGAGATGAA	qPCR for *OsNKD1*
OsNKD1-qPCR-R	ATGCGTTAATCCCCCAGAGT	
OsDEK1-qPCR-F	GACTGAGCGAATGAAGCACAA	qPCR for *OsDEK1*
OsDEK1-qPCR-R	AACGCATCTCGGGTGGA	

Table S5. Summary of WGBS technical data generated in this study.

Samples	Clean reads	Mapped reads	Unique mapped reads	Conversion rate (%)	Mean coverage depth (x)
ZH11 embryo	66,885,332	54,624,542	47,913,246	99.39	18.59
ZH11 endosperm	72,549,332	60,363,538	54,848,450	99.52	21.38
ta2-1 endosperm	67,335,332	54,697,576	47,230,964	99.61	18.37

Dataset S1. List of hypermethylated and hypomethylated DMRs in *ta2-1*.