the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2022) **109,** 1035–1047 doi: 10.1111/tpj.15698

Temporal changes in transcripts of miniature inverted-repeat transposable elements during rice endosperm development

Hiroki Nagata[1], Akemi Ono[1], Kaoru Tonosaki[1,2], Taiji Kawakatsu[3], Yutaka Sato[4], Kentaro Yano[5], Yuji Kishima[6] and Tetsu Kinoshita[1]*

[1]*Kihara Institute for Biological Research, Yokohama City University, 641-12 Maioka, Totsuka, Yokohama, Kanagawa 244-0813, Japan,*

[2]*Faculty of Agriculture, Iwate University, 3-18-8 Ueda, Morioka, Iwate 020-8550, Japan,*

[3]*Institute of Agrobiological Sciences, National Agriculture and Food Research Organization, 3-1-3 Kan-nondai, Tsukuba, Ibaraki 305-8604, Japan,*

[4]*Genetic Strains Research Center, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan,*

[5]*Department of Life Sciences, School of Agriculture, Meiji University, 1-1-1 Higashi-mita, Kawasaki 214-8571, Japan, and*

[6]*Research Faculty of Agriculture, Hokkaido University, Kita-9 Nishi-9, Kita-ku, Sapporo 060-8589, Japan*

Received 12 October 2021; revised 19 January 2022; accepted 27 January 2022.; published online 2 February 2022.
*For correspondence (e-mail tkinoshi@yokohama-cu.ac.jp).

SUMMARY

The repression of transcription from transposable elements (TEs) by DNA methylation is necessary to maintain genome integrity and prevent harmful mutations. However, under certain circumstances, TEs may escape from the host defense system and reactivate their transcription. In Arabidopsis (*Arabidopsis thaliana*) and rice (*Oryza sativa*), DNA demethylases target the sequences derived from TEs in the central cell, the progenitor cell for the endosperm in the female gametophyte. Genome-wide DNA demethylation is also observed in the endosperm after fertilization. In the present study, we used a custom microarray to survey the transcripts generated from TEs during rice endosperm development and at selected time points in the embryo as a control. The expression patterns of TE transcripts are dynamically up- and downregulated during endosperm development, especially those of miniature inverted-repeat TEs (MITEs). Some TE transcripts were directionally controlled, whereas the other DNA transposons and retrotransposons were not. We also discovered the NUCLEAR FACTOR Y binding motif, CCAAT, in the region near the 5′ terminal inverted repeat of *Youren*, one of the transcribed MITEs in the endosperm. Our results uncover dynamic changes in TE activity during endosperm development in rice.

Keywords: transposable element, MITE, endosperm, *Oryza sativa*, CCAAT motif.

Linked article: This paper is the subject of a Research Highlight article. To view this Research Highlight article visit https://doi.org/10.1111/tpj.15711

INTRODUCTION

The transcription of autonomous transposable elements (TEs) is essential for the transposition of class I RNA retrotransposons and class II DNA transposons (Wells & Feschotte, 2020) but is typically silenced by epigenetic mechanisms such as DNA methylation (Zhang et al., 2018). DNA methylation also disarms non-autonomous elements via transcriptional repression of their RNA intermediate or by blocking transposase binding to its target sequence required for 'cut-and-paste' activity (Gierl et al., 1988; Hashida et al., 2006). Although DNA methylation plays essential roles in the repression of TE activity, this epigenetic mark is passively removed from TE sequences during DNA replication or actively erased by DNA glycosylases during the plant life cycle and in various tissues, including seed endosperm (Bartels et al., 2018; Ono & Kinoshita, 2021).

The endosperm is the nutritive tissue that sustains embryonic growth during seed development and seedling growth after germination, especially in crop species such as rice (*Oryza sativa*) (Lopes & Larkins, 1993; Sabelli & Larkins, 2009). The endosperm arises from the second fertilization event between the central cell of the female gametophyte and one of the two sperm cells delivered by the male pollen tube (Berger et al., 2008). The embryo is

the product of the fertilization of the egg and the other sperm cell. Unlike the embryo, the endosperm has unique characteristics; its genome comprises two maternal copies and one paternal copy, it undergoes terminal differentiation, and it is subjected to global DNA demethylation (Hsieh et al., 2009; Lauria et al., 2004; Zemach et al., 2010).

In Arabidopsis (*Arabidopsis thaliana*), a DNA glycosylase termed DEMETER targets AT-rich transposons in euchromatin (Gehring et al., 2006; Ibarra et al., 2012), as well as the heterochromatic TEs with the histone chaperone FACT in the central cell, the progenitor cell of the endosperm before fertilization (Frost et al., 2018; Ikeda et al., 2011). This active demethylation also contributes to parent-of-origin-specific gene expression, known as imprinting, in the endosperm (Fujimoto et al., 2011; Hsieh et al., 2011), possibly via an association with TEs (Qiu & Kohler, 2020). Indeed, many TEs are assumed to be activated in the endosperm, which is consistent with pioneering work exploring the causes of kernel pigmentation patterns (Doring & Starlinger, 1986) and maternal DNA hypomethylation in maize (*Zea mays*) endosperm (Lauria et al., 2004). The activity of TEs is typically repressed by their host defense system to prevent harmful mutations and maintain genome integrity; the global DNA demethylation observed in the endosperm constitutes a notable exception during plant development. The current explanation predicts that TE transcripts from the central cell or the endosperm are processed into small RNAs, which are then transported to the egg cell and the embryo to reinforce the repression of TEs in their genomes for the next generation (Bourc'his & Voinnet, 2010). In addition, the endosperm is a terminally differentiated tissue and might therefore tolerate high TE activity brought upon by DNA demethylation. De-repression of TE activity influences the neighboring genomic regions (Naito et al., 2009; Yang et al., 2005), making the endosperm an ideal tissue to dissect the roles of TEs in relation to transcriptional control and genome integrity.

The rice genome is relatively small (389 Mb) compared to those of other crop species and only approximately three times larger than that of Arabidopsis (125 Mb) (Arabidopsis Genome Initiative, 2000; International Rice Genome Sequencing, 2005). However, rice contains many more repetitive sequences than does Arabidopsis (Song & Cao, 2017). The TE distribution is also different: TEs in Arabidopsis are mainly located in the pericentromeric regions, whereas those in rice also populate the euchromatin (Mirouze & Vitte, 2014). A striking feature of the rice genome with regard to copy number and TE distribution is its enrichment in DNA transposons, especially miniature inverted-repeat TEs (MITEs) (Jiang et al., 2004). MITEs are internal deletion derivatives of autonomous DNA transposons with terminal inverted repeats (TIRs) at both ends (Feng, 2003; Feschotte et al., 2002; Feschotte & Wessler, 2002). These short stretches of sequences, 100–600 bp in length, mostly reside in the euchromatic regions and are relatively close to genes compared to longer long terminal repeat-type retrotransposons (Oki et al., 2008; Zemach et al., 2010).

In the present study, we performed a time-course transcriptome analysis of rice TEs during endosperm development. We employed a custom microarray designed to detect transcripts generated from repetitive sequences in rice (Ishiguro et al., 2014), aiming to bypass the limitations associated with short-read sequencing of TEs (Oshlack & Wakefield, 2009), especially for short and high-copy-number elements such as MITEs. We determined that many expressed TEs in the endosperm tended not to be expressed in the embryo and showed a time-dependent expression pattern. Among the expressed TEs, many MITEs displayed directional behavior because we detected transcripts originating from one strand only, probably reflecting the direction of the ancestral autonomous transcriptional unit. This pattern was not present in other tested DNA transposons and retrotransposons. We also identified the binding motif for NUCLEAR FACTOR Y (NF-Y) transcription factors (Mantovani, 1999; Nardini et al., 2013) within *Youren* sequences, which overlapped with the boundary of the 5′ TIR and the internal region, and provide supporting evidence that this motif promotes transcription of *Youren*. Taken together, our results suggest that, although MITEs are inserted proximal to genes or coding regions in random orientations, MITE transcripts do not hitchhike with the directional transcription of their neighboring genes (i.e. readthrough or chimeric transcripts) but rather exhibit their own directional transcriptional control, possibly starting from intrinsic sequences in *Youren* MITEs. Such an active transcriptional environment at MITE sequences may contribute to shaping the transcriptional network of genes expressed in the rice endosperm.

RESULTS

Expression of TEs in the embryo and the endosperm

Genome-wide hypomethylation and TE transposition are prominent features of the endosperm, but how TE expression changes over time during seed development is not well described. Therefore, we conducted a transcriptome analysis with a custom 44K Agilent microarray (Ishiguro et al., 2014) harboring 31 366 repeat-related probes, 864 microRNA (miRNA)-related probes, and probes for 9563 genes. The probes were designed from annotated TEs in the TIGR Plant Repeat Database (Ouyang & Buell, 2004). To investigate the dynamics of TE transcription, we isolated total RNAs from rice endosperm in 24-h intervals from 2 to 7 days after pollination (DAP) and at 10 DAP (Figure S1a). The endosperm became coenocytic at 2 DAP and grew in a centripetal direction by 3 DAP. Cellularization took place between these time points under our growth conditions (Tonosaki et al., 2021). The endosperm

continued to divide and began to deposit storage compounds until the maturation stage at 10 DAP (Figure S1a). Our chosen time points thus covered most developmental transitions in endosperm development. In addition to the endosperm, we also extracted total RNAs from developing embryos collected at 5, 7, and 10 DAP as controls, which correspond to the coleoptile, the completion of organogenesis, and the maturation stages of rice embryogenesis, respectively (Itoh et al., 2016) (Figure S1b).

We performed a principal component analysis using the expression of genes and TEs obtained from all samples across replicates. We observed separation of genes and TEs expressed in the embryo from those expressed in the endosperm, as well as high reproducibility across biological replicates (Figure S1c–e). In addition, marker genes for embryo or endosperm exhibited the expected expression patterns (Figure S1f). Using matched sample pairs from embryo and endosperm collected at 5, 7, and 10 DAP, we drew MA plots to explore the underlying expression differences for TEs. We detected 32 (5 DAP), 435 (7 DAP), and 506 (10 DAP) differentially expressed (DE) TEs during embryogenesis and 105 (5 DAP), 554 (7 DAP), and 2123 (10 DAP) DE TEs during endosperm development (Figure 1a–c). TE expression may be suppressed in the embryo if it is high in the endosperm, according to the proposed hypothesis that siRNAs reinforce TE silencing in the embryo (Bourc'his & Voinnet, 2010). To test this possibility, we compared the expression levels of DE TEs that were 'upregulated in embryo' and 'upregulated in endosperm' using rain cloud plots (Figure 1d–f). Although we observed many upregulated TEs in the embryo in our matched sample pairs at 5, 7, and 10 DAP, many upregulated TEs in the endosperm showed a correspondingly low expression level in the embryo (Figure 1d–f). These results were consistent with the proposed hypothesis that TEs are highly transcribed to silence their counterparts in the embryo genome, although it is also possible that TE expression patterns merely reflect the differences in the developmental programs of the embryo and endosperm. In any case, we observed that the expression patterns of TEs were largely distinct between the embryo and the endosperm, as also seen for protein-coding genes (Figure 1a–c; Figure S1f).

Temporal expression patterns of TEs during rice endosperm development

We focused on all samples derived from the endosperm to investigate the temporal expression patterns of the repetitive sequences, including those of TEs, during rice endosperm development from 2 to 10 DAP. We selected DE probes including retrotransposons, DNA transposons, MITEs, repeats, and genes (Tables S1 and S2) by one-way ANOVA and Storey methods (see Materials and methods). Then, we detected dynamic and stage-specific expression patterns using a *k*-means clustering approach, which produced 10 distinct clusters (Table S3). Clusters A3, A6, and A7 (Figure 2a) showed an upregulation at 2 and 3 DAP, which may reflect DNA demethylation of the central cell and early endosperm development. However, most repeats were upregulated at later stages, between 7 and 10 DAP, as illustrated by clusters A2, A4, A8, A9, and A10 (Figure 2a). This result was consistent with the pairwise comparisons of the embryo and endosperm samples at 10 DAP, at which stage more TEs were activated in the endosperm (Figure 1c,f). Among the DE TEs during endosperm development, the MITE family was prominent in each cluster relative to retrotransposons, DNA transposons or other repeats (Figure 2b), which might reflect the higher copy number of MITEs in the rice genome. Our analysis also revealed enrichment of retrotransposons in DE probes against designed probes in the microarray analysis (Table S2). However, in the present study, we focused on shorter transcripts of MITEs taking advantage of microarray analysis. The standard RNA-sequencing protocols underestimate relatively short RNAs (Oshlack & Wakefield, 2009), which are fragmented, primed with random hexamers, and then recovered by the bead technology, resulting in the loss of relatively short RNAs such as snRNA, snoRNA, tRNA, and intrinsic transcripts from MITEs during library preparation.

We extracted the data from MITE probes only for reanalysis (Figure 3a; Table S4) and sorted this into individual MITE subfamilies (Figure 3b). We assessed detectable expression for the MITE subfamilies *Tourist*, *Castaway*, *MITE-adh type M*, *Gaijin/Gaigin*, and *Ditto*. These MITE subfamilies were differentially expressed during endosperm development, with broad expression patterns, because all clusters contained representative members of all subfamilies. By contrast, a set of probes corresponding to *Youren* MITEs was predominantly enriched in the M6 cluster (Figure 3c). *Youren* is a relatively short non-autonomous TE (approximately 250–300 bp) (Figure S2a) belonging to the *Tourist*-like elements with 3-bp (TAA or TTA) target site duplications (Zhang et al., 2004). These elements were characterized by one of the highest signal intensities among all DE TEs in the endosperm at 5 DAP (Figure 3d), whereas signal intensities were low in the embryo, anthers, and leaves (Figure S2b).

To assess the possibility of cross-hybridization of highly conserved repetitive transcripts such as those derived from TEs, we investigated the relationship between the expression levels and the level of identity between probes (Figure 4a,b). This possibility was of special concern for *Youren* elements because they showed overlapping expression patterns but were limited to a single cluster (Figure 3a,b). We therefore subdivided all *Youren* elements into 11 groups based on their expression level (Figure 4a) and generated a phylogenetic tree of each probe recognizing each *Youren* element, which revealed no correlation

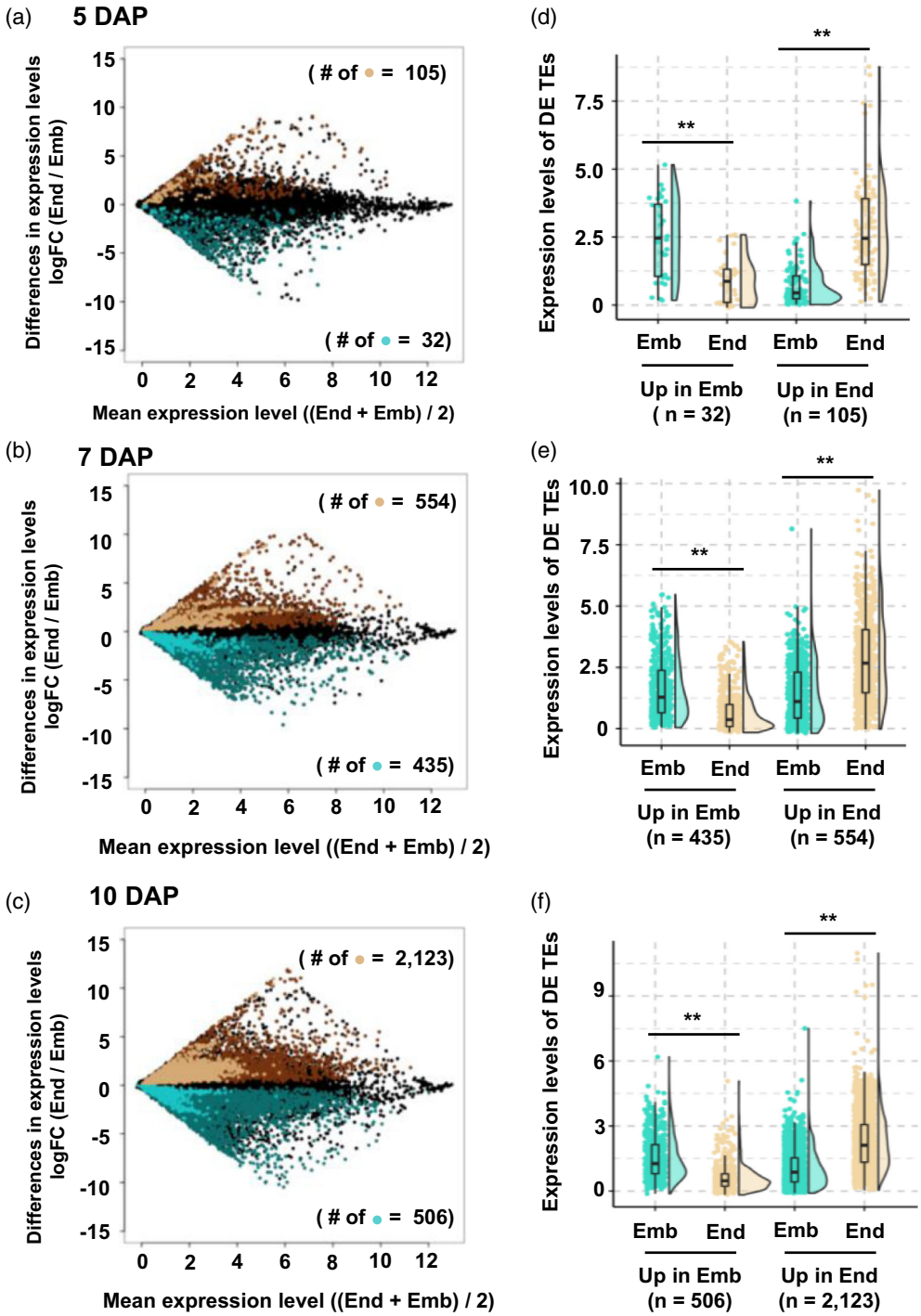

Figure 1. Expression patterns of genes and TEs in the rice embryo and endosperm.
(a–c) MA plots showing the expression levels of genes and TEs at 5 DAP (a), 7 DAP (b), and 10 DAP (c). The light and dark brown dots represent differentially DE TEs and genes upregulated in the endosperm. The light and dark green dots represent DE TEs and genes upregulated in the embryo. The black dots indicate non-significant probes of TEs and genes (FDR < 0.001). (d–f) Rain cloud plots showing the expression levels of DE TEs. 'Up in Emb' and 'Up in End' represent negative and positive $\log_{FC}$ dots from MA plot data. The light blue and light brown dots indicate individual expression levels in the embryo and endosperm fractions. *$P < 0.05$; **$P < 0.001$ using a Wilcoxon rank sum test.

between sequence similarity and signal intensity except for the 10th and 11th weakly expressed groups (Figure 4b). Therefore, although we cannot completely exclude the potential for cross-hybridization in the microarray analysis, we know that those transcripts were not generated from single loci in the rice genome.

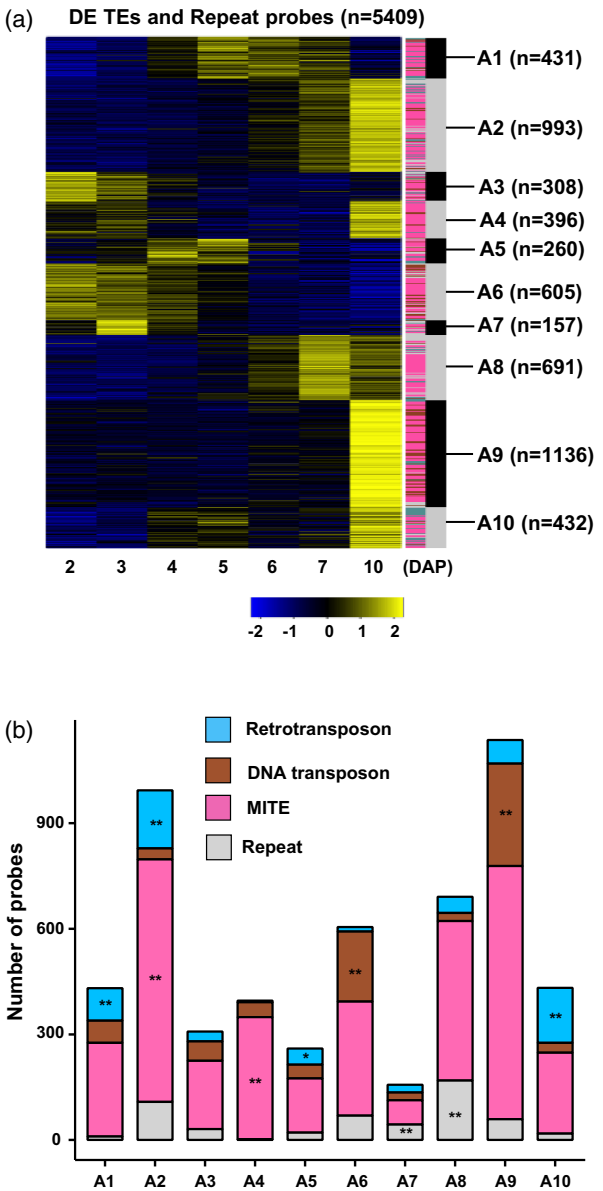

Figure 2. Temporal expression patterns of repetitive sequences.
(a) Heatmap representation of *k*-means clustering of the row *Z*-scored expression patterns of DE TEs and repeats during endosperm development. The colored lines to the right denote representative TE and repeat subclasses: retrotransposons, magenta; DNA transposons, orange; MITEs, deep green; and repeats, gray. The gray and black bars indicate individual clusters. (b) Number of DE probes for the representative subclasses of TEs and repeats in each *k*-means cluster. Significantly enriched DE probes of individual TE subfamilies in each cluster against all DE TE probes are indicated (*$Q < 0.05$; **$Q < 0.01$). *Q*-values were calculated using a hypergeometric test adjusted using the Benjamini–Hochberg method.

Because MITE sequences are deposited in genome databases without regard for their orientation (Ouyang & Buell, 2004), reference MITE sequences comprise a mixture of forward and reverse orientations (Zhang et al., 2004). A phylogenetic tree built from full-length *Youren* sequences in the databases formed two distinct clades (Figure 4c). To determine the direction of *Youren* transcripts, we reannotated each *Youren* probe as reference and reverse orientation, according to the orientation of each *Youren* element deduced from the deposited sequence, taking the direction of transcripts from the putative autonomous copy as a guide. The 60-mer probes printed onto the microarrays were designed to hybridize to the corresponding 5′ and 3′ end sequences of annotated *Youren* elements, making it possible for probes to detect antisense and sense transcripts (Ishiguro et al., 2014) (Figure 4d). Most *Youren* transcripts appeared to hybridize to the 3′ probes, suggesting that their transcripts are in the reference orientation (Figure 4d,e). We confirmed that *Youren* transcripts are transcribed from the reference strand by a reverse transcriptase-polymerase chain reaction (RT-PCR) (Figure S3). In addition, we detected signals using the antisense probe but not the sense probe in transverse sections of 4 DAP endosperm by *in situ* hybridization (Figure S4). Thus, our results suggest the coordinated expression of *Youren* elements originating from different copies scattered over the rice genome. A similar trend of temporal expression patterns was associated with the direction of transcription for *Tourist*, *Castaway*, *MITE-adh type M*, and *Gaijin* MITE elements, but not for *Ditto* (Figure S5).

Identification of the NF-Y binding motif in *Youren*

We aimed to understand the basis for the directional transcription of MITEs. Although we could not distinguish individual insertions of multicopy MITE sequences from the microarray data, we determined locations of all copies of DE TEs and non-DE TEs in the rice genome. The resulting karyoplots indicated that all tested MITE subfamilies map to all chromosomal arm regions, in agreement with a previous report (Mirouze & Vitte, 2014), for both DE MITEs and non-DE MITEs (Figure S6). We again focused on highly expressed *Youren* (we used 198 different DE *Youren* copies because eight copies were detected by probes hybridizing to both ends, pointing to these copies as duplications). We also failed to detect any significant skew between MITE subfamilies and DE and non-DE *Youren* elements for the position of their insertion sites relative to the upstream, coding, intron or downstream regions of protein-coding genes (Figure 5a). The orientations of *Youren* insertions relative to the direction of transcription for their closest gene appeared to be random (Figure S7). We thus concluded that these directional transcripts likely arose from internal *Youren* sequences.

Because transcript directionality can be attributed to the presence of *cis*-regulatory elements in 5′ sequences, we searched for such elements in all *Youren* copies. A CCAAT motif, which is the typical binding site for NF-Y transcription factors, was present at the boundary between the 5′ TIR and the internal sequence of *Youren*. We also

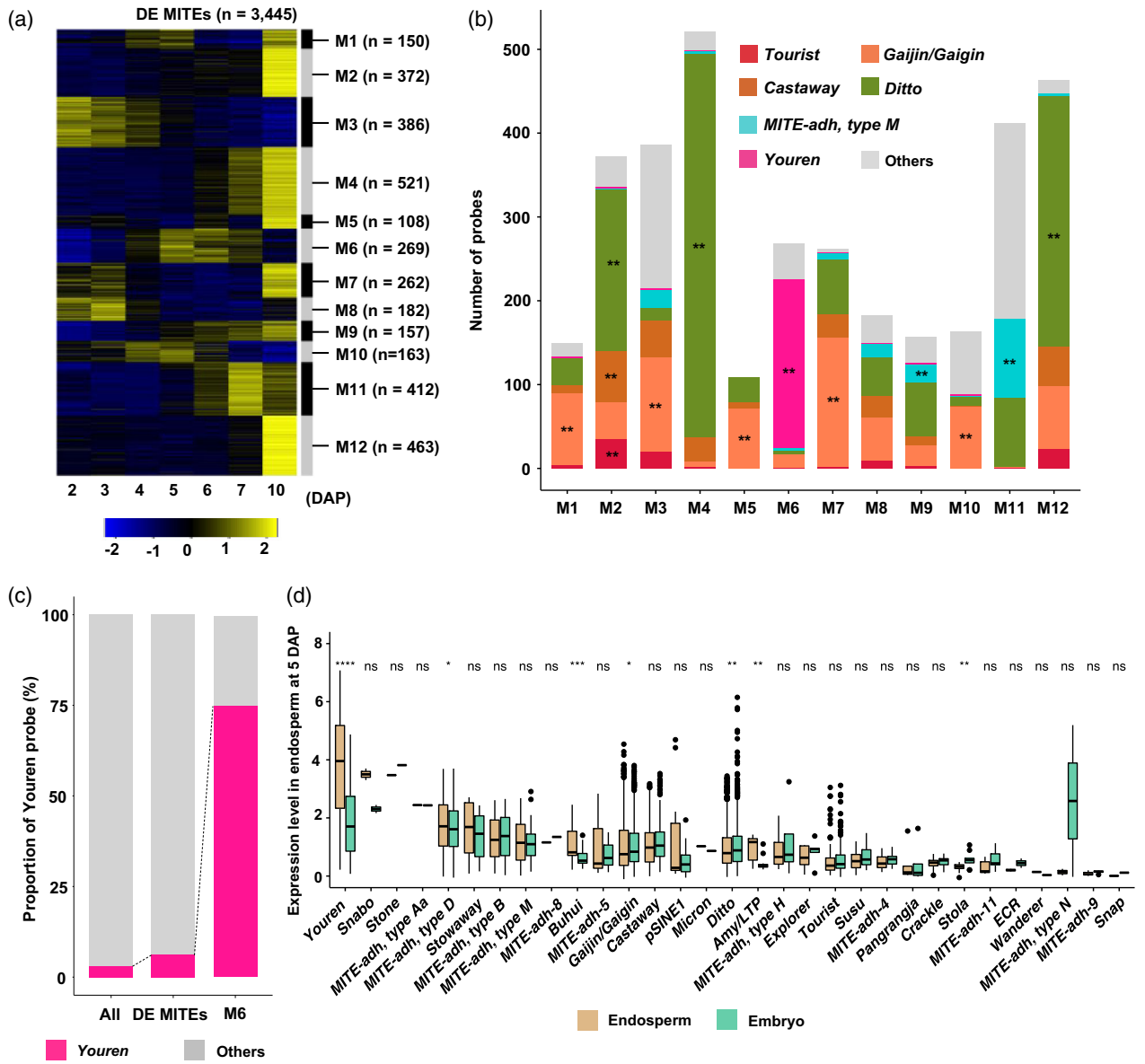

Figure 3. Temporal expression patterns of MITE subfamilies during rice endosperm development.
(a) Heatmap representation of row *Z*-scored expression patterns of DE MITEs during endosperm development. Gray and black bars indicate each cluster from *k*-means analysis. (b) Number of probes for each representative DE MITE subfamily in individual clusters. The statistical significance of the enrichment is indicated (*$Q < 0.05$; **$Q < 0.01$) for the number of DE probes of individual TE subfamilies in each cluster against all DE TE probes. *Q* values were calculated using a hypergeometric test adjusted using the Benjamini–Hochberg method. (c) Enrichment of the *Youren* subfamily in the M6 cluster among all designed MITE and DE MITE probes. *Youren* probes, magenta; other MITE subfamilies, gray. (d) Box plots of $\log_2$-transformed normalized signal intensity for each MITE subfamily in embryo (green) and endosperm (brown) at 5 DAP. At least one probe for each individual subfamily was differentially expressed in the endosperm from 2 to 10 DAP. Box plots were sorted using the $\log_2$-transformed normalized signal intensity of endosperm at 5 DAP. The *Youren* subfamily shows one of the highest expression levels. Statistical significance (*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$; ****$P < 0.0001$; ns, not significant) was calculated using a Wilcoxon rank sum test.

identified a fragment of this motif in the 3′ TIR as a consequence of the inverted repeat, although the internal sequence did not provide the necessary sequence to reconstitute the entire motif (Figure 5b). The motif probably enhances transcription of the *Youren* copies: the expression levels of the copies with the motif were relatively higher than those of the copies without the motif (Figure 5c,d). We extended our *in silico* analysis to the entire *Youren* sequences using the online tool Plant Promoter Analysis Navigator (PLANTPAN 3.0; http://plantpan.itps.ncku.edu.tw), which identified Homeodomain TALE, GT1CONSENSUS, and AT-Hook *cis*-elements (Boulikas, 1995;

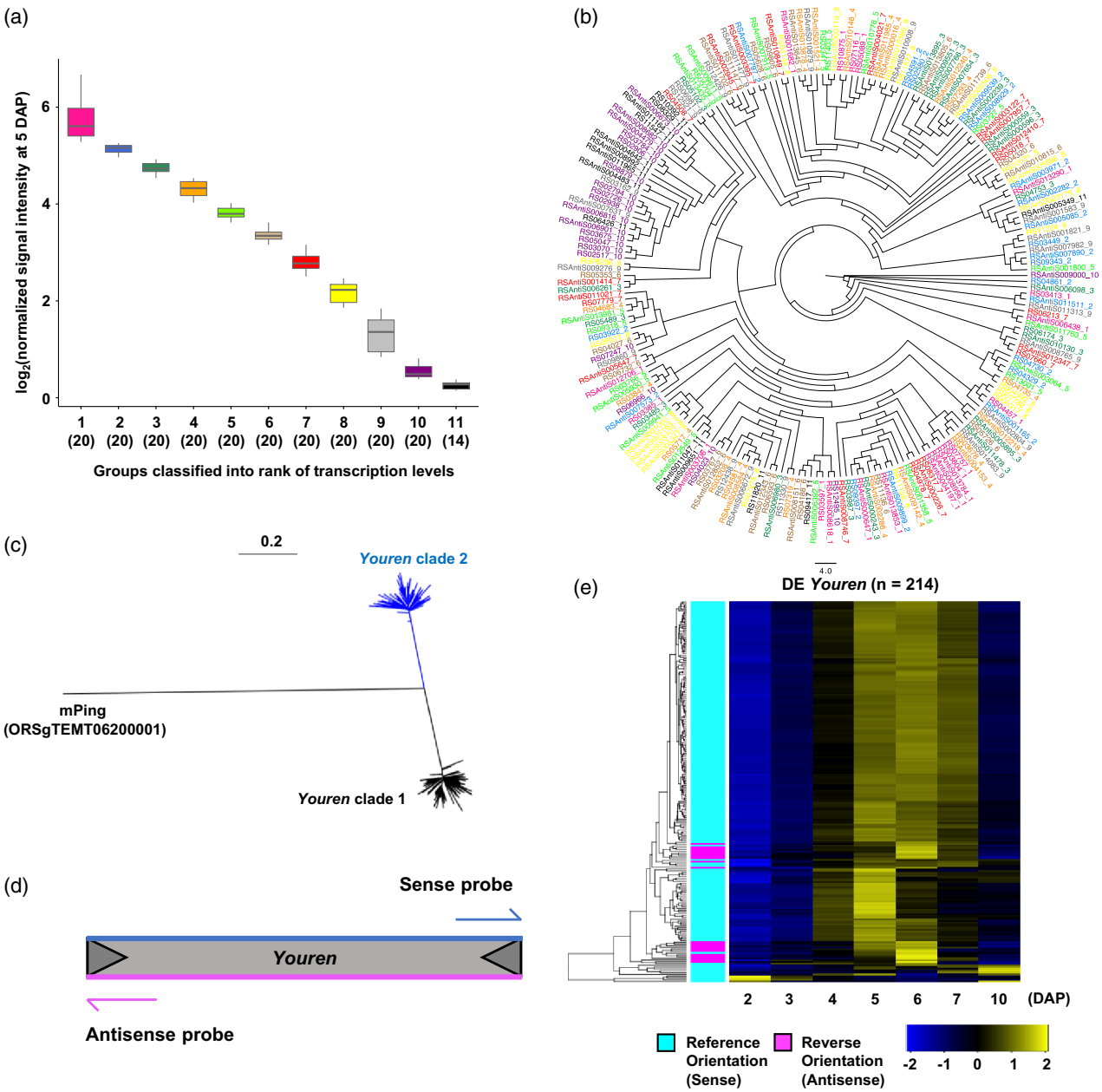

Figure 4. *Youren* TEs are preferentially transcribed in a single direction that matches that of the putative autonomous copy.
(a) *Youren* expression levels, arbitrarily divided into 11 groups (214 in total) based on $\log_2$-transformed normalized signal intensity in the endosperm at 5 DAP. Numbers in parentheses indicate the number of probes in each group. (b) Sequence similarity of 214 DE *Youren* probes (60-mer), shown as a diagram made using CLUSTALW and FIGTREE. (c) Phylogenetic tree illustrating the sequence similarity of the entire *Youren* region. Clades 1 and 2 are composed of complementary sequences. *mPing*, a member of the *Tourist*-like subfamily of MITEs, was used for the outgroup, together with *Youren*. (d) Schematic representation of *Youren* TEs and the position of the microarray probes. The Cy-3-labeled complementary RNA is synthesized in reverse direction to the *Youren* transcripts. Therefore, the sense probe detects transcripts in reference orientation (sense strand), whereas the antisense probe detects transcripts in reverse orientation (antisense strand). (e) Heatmap representation of the relationship between expression patterns and transcriptional direction. Color key indicates row *Z*-score of $\log_2$-transformed normalized signal intensity. Cyan, *Youren* in reference orientation; magenta, *Youren* in reverse orientation.

Hamant & Pautot, 2010; Kyozuka et al., 1994; Xu et al., 2013). These transcription factor binding sites were preferentially located in the 5′ region of *Youren* for forward orientation, but not for reverse orientation of *Youren* (Figure 5e) or other tested MITEs. For *Youren* elements with detectable transcripts by RT-PCR, the AT-Hook binding sites were located within the forward primer used for PCR, suggesting that the element overlaps with the transcribed *Youren* region (Figure S3). The NF-Y binding motif is a good candidate to explain the directional transcription of

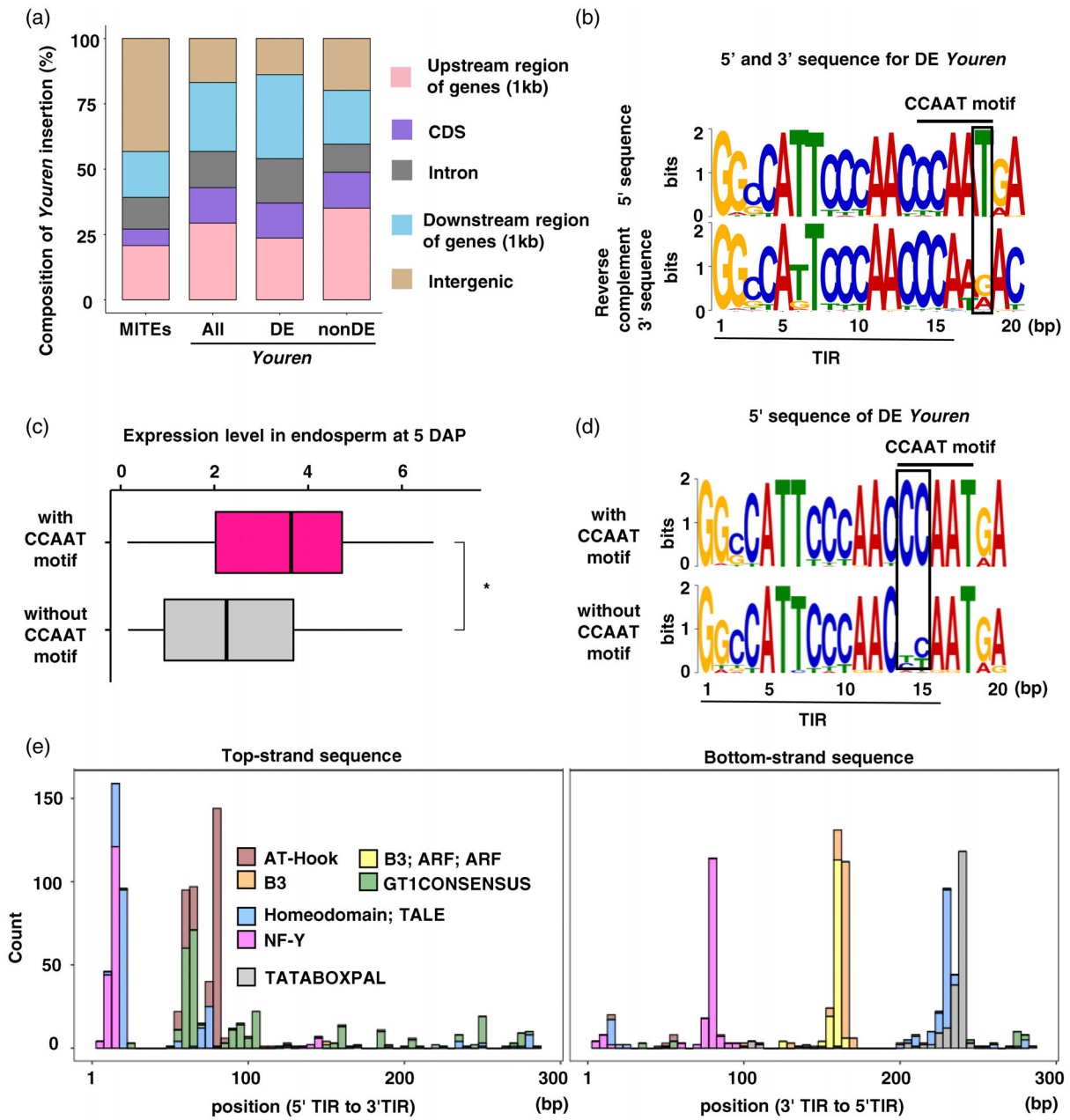

Figure 5. *Youren* insertions in the genome and transcription factor binding motifs.
(a) Proportion of insertion sites relative to genes for all MITEs, all *Youren* sequences, as well as DE and non-DE *Youren* sequences. (b) Sequence logos for the 5′ (upper) and 3′ reverse complementary (bottom) sequences that include the TIR and the inner regions, drawn using the MEME suite (https://meme-suite.org). The horizontal line above the logos indicates the conserved CCAAT motif, known as the NF-Y binding site. The box indicates the last T of the CCAAT motif, which is not conserved in the 3′ sequences of DE *Youren*. (c) Box plots of the expression level of *Youren* with (magenta) and without (gray) the CCAAT motif in endosperm at 5 DAP. *$P < 0.05$ using a Wilcoxon rank sum test. (d) Sequence logos for the 5′sequences of *Youren* with (upper, $n = 165$) and without (bottom, $n = 33$) the CCAAT motif. The horizontal line above the logos indicates the conserved CCAAT motif. The box indicates the initial CC of the CCAAT motif, which is not conserved in the minor group of *Youren* sequences. (e) Stacked histogram showing the positions of representative transcription factor binding sites and the top strand (left) and bottom strand (right) of DE *Youren*. *In silico* analysis of transcription factor binding sites of DE *Youren* sequences ($n = 198$) was performed with PLANTPAN 3.0 (http://plantpan.itps.ncku.edu.tw).

Youren because we detected expression of *NF-YA2*, *NF-YA4*, *NF-YA6*, and *NF-YA9* in the embryo and endosperm in our microarray analysis (Figure S8). In addition, although probes related to the other trimeric components, *NF-YB* and *NF-YC*, are not printed on the microarray, several genes for *NF-YB* (i.e. *NF-YB1*, *NF-YB7*, and *NF-YB9*)

are predominantly expressed in the aleurone and starchy endosperm and control seed development in rice (Bai et al., 2016; Niu et al., 2020, 2021; Zhigou et al., 2018). A comprehensive approach for understanding gene expression patterns and interactions of each NF-Y component has been taken, including in rice endosperm tissue (Zhigou et al., 2018). In conclusion, we identified coordinated expression patterns of *Youren* and the potential underlying *cis*-elements.

DISCUSSION

Hypomethylation of the genome is observed in the endosperm of several plant species, which prompted us to investigate the expression patterns of rice TEs during endosperm development and the epigenetic control mechanisms of TE expression in relation to protein-coding gene expression and genome integrity. TEs can influence the expression of their neighboring genes, either by preventing transcription or by promoting transcription via providing an auspicious chromatin environment for active transcription (Feschotte, 2008; Qiu & Kohler, 2020). In addition, some transcripts derived from TEs exhibit a sponge-like activity against small RNAs and miRNAs, hybridizing and sequestering them away from their targets (Cho & Paszkowski, 2017). In other cases, transcripts from MITEs provide the stem structure required to form the 'stem-loop' structure of pre-miRNAs. Therefore, temporal analysis of TE expression originating from hypomethylated endosperm is an ideal resource to study the roles of TEs.

Because genome-wide demethylation has been observed in the central cell and endosperm cells, a simple prediction is that TE activity is high in the endosperm at all developmental stages (Ibarra et al., 2012; Park et al., 2016). However, our temporal analysis revealed a very different picture because the expression of TEs was not static and high but exhibited dynamic changes, with some TEs expressed during earlier stages and others expressed later. We observed a similar trend in the embryo, although we limited our investigation to three time points, representing the coleoptile, the completion of organogenesis, and the maturation stages of rice embryogenesis. Over the course of these time points, the embryo develops into a mature structure that consists of several tissues: the root and shoot primordium, leaves, and the scutellum. The expression of rice TEs is largely different in distinct plant tissues (Cho & Paszkowski, 2017); our results therefore probably merely reflect the tissue- and organ-specific transcriptional programs in the embryo. Similarly, the rice endosperm develops into several domains, such as the peripheral region and the region adjacent to the embryo, as well as the aleurone layer and the starchy endosperm, and the expression profile of TEs may differ in each compartment. Alternatively, the transcriptional environment of TEs may rely on temporal programs and epigenetic changes of TE sequences. Demethylation of the maternal genome occurs in several plant species, resulting in imprinted gene expression (Batista & Kohler, 2020; Hsieh et al., 2009; Rodrigues et al., 2013). MITEs including *Youren* have the potential to contribute to this type of epigenetic modulation of gene expression because maternal demethylation is enriched in those subfamilies (Yuan et al., 2017). In addition to the demethylation of the maternal copy of the endosperm genome, DNA methylation status changes during seed and endosperm development in both Arabidopsis and rice (Kawakatsu et al., 2017; Xing et al., 2015). To better understand the roles of TEs and demethylation in gene expression during endosperm development, it is necessary to conduct higher-resolution expression analyses at the cell-type level, combined with multiple omics analyses to uncover the epigenetic states underlying endosperm development.

Although the transcription of long and almost intact TEs such as long terminal repeat retrotransposons is directionally controlled, the transcription of remnant TEs is assumed to be influenced by the expression of neighboring genes and may even form chimeric transcripts bringing together gene sequences and TEs. MITEs are remnant TEs derived from longer autonomous copies and thus no longer encode the sequences required for their transposition. Nevertheless, transcripts derived from MITE loci, with the exception of *Ditto*, are directionally controlled. Our results suggest that these elements might have preserved the promoter activity required for autonomous transcription. Among the tested MITEs, *Youren* was unique because the expression of most elements was relatively high and similarly and directionally controlled. We hypothesize that these patterns are caused by *cis*-elements that reside in the 5′ internal region of *Youren* because the sequences outside of TE insertion sites are not conserved and cannot explain their coordinated expression.

The approach that we followed in the present study provides the framework to test the proposed hypothesis: TEs provide *cis*-elements that can therefore rewire transcriptional networks (Feschotte, 2008). Although direct binding and transcriptional control will need to be tested, the trimeric NF-Y complex is a good candidate. The NF-YB and NF-YC subunits have histone-fold domains to bind to DNA in a non-sequence-specific manner, whereas NF-YA recognizes the CCAAT motif (Nardini et al., 2013). Although pioneer transcription factors are not known in the endosperm, Arabidopsis LEAFY COTYLEDON1 (LEC1), an NF-YB subunit, is involved in embryogenesis (Tao et al., 2017). We identified 198 *Youren* copies actively transcribed during the mid-stage of endosperm development, the timing of which is consistent with the onset of transcription for genes encoding starch synthases and storage proteins (Tonosaki et al., 2021). These copies of *Youren* are inserted in 700 distinct locations in the rice genome close to genes and

provide potential *cis*-elements for trimeric NF-Y binding. Our work provides useful information about the intrinsic transcription of MITEs, which tends to be lost when using standard RNA-sequencing protocols. Based on our results, it will be worth investigating whether and how much these elements can rewire transcriptional networks in the endosperm in the context of the evolutionary history of rice.

EXPERIMENTAL PROCEDURES

Plant materials and growth conditions

Rice plants (*Oryza sativa* cv. Nipponbare) were grown in a Biotron chamber (NC-350HC; Nippon Medical & Chemical Instruments Co. Ltd, Osaka, Japan) in accordance with a previously described protocol (Ohnishi et al., 2011) adjusted for short-day photoperiod and temperature cycles (11 h of light at 30°C/13 h of dark at 25°C).

RNA extraction from rice endosperm

Rice endosperm was harvested at 24-h intervals from developing seeds from 2 to 7 DAP and at 10 DAP (Figure S1a). For the isolation of liquid coenocytic endosperm (2 DAP) and just after cellularization (3 DAP), a cut was made on the surface of the seed coat using a scalpel under a binocular microscope, as described previously (Tonosaki et al., 2021). The liquid endosperm was collected from the cut by pressing on the seed using a pipette and added to 50 μl of extraction buffer from the PicoPure RNA isolation kit (Applied Biosystems, Waltham, MA, USA). Between 70 and 100 seeds at 2 DAP and 50 seeds at 3 DAP were used for isolation of liquid endosperm. After 4 DAP, endosperm tissues were directly dissected from the ovaries under a binocular microscope, followed by homogenization in RNA isolation cocktail, consisting of 350 μl of RLC buffer (Plant RNeasy kit; Qiagen, Hilden, Germany), 100 μl of Fruit-mate (Takara, Shiga, Japan), and 4.5 μl of β-mercaptoethanol (Wako, Osaka, Japan) (Lee et al., 2016) using micro pestles and 1.5-ml plastic tubes. Between 10 and 30 ovaries were used for isolation of endosperm from 4 to 7 DAP. Total RNA was extracted with the RNeasy Plant Mini Kit in accordance with the manufacturer's instructions (Qiagen). At 10 DAP, dissected endosperm tissue from three ovaries was directly transferred to 1.5-ml plastic tubes and immediately frozen in liquid nitrogen. The frozen endosperm was homogenized with a multibead shocker (MB755U; Yasui Kikai, Osaka, Japan) at 1500 r.p.m. for 2 sec. The resulting homogenized endosperm was resuspended in 1.3 ml of Fruit-mate. The mixture was subdivided into two 650-μl aliquots and centrifuged at 12 000 ***g*** for 5 min at 4°C. To 250 μl of each supernatant, 400 μl of RLT buffer from the RNeasy Plant Mini Kit (Qiagen) was added together with 4 μl of β-mercaptoethanol (Wako). For DNase treatment, RNase-Free DNase Set (Qiagen) was used in accordance with the manufacturer's instructions.

RNA extraction from rice embryos

Rice embryos were dissected from developing ovaries at 5, 7, and 10 DAP (Figure S1b) under a binocular microscope. Because early-stage embryos are small and difficult to dissect under a binocular microscope, two protocols were used to isolate RNA from embryos at 5 DAP: the RNeasy Plant Mini Kit and the PicoPure RNA isolation kit (Thermo Fisher Science, Waltham, MA, USA). For the RNeasy Plant Mini Kit, 15 embryos were collected from developing ovaries and were immediately frozen in liquid nitrogen. The frozen embryos were homogenized with a multibead shocker at 1500 r.p.m. for 2 sec; the subsequent steps followed the protocol described above. For the PicoPure RNA isolation kit (Thermo Fisher Science), three embryos at 5 DAP were homogenized in 10 μl of extraction buffer from the PicoPure RNA isolation kit using a micro pestle and a 1.5-ml microcentrifuge tube. The homogenate was incubated at 42°C for 30 min. Total RNA was extracted in accordance with the manufacturer's instructions. At 7 and 10 DAP, three embryos were collected for extraction of total RNA using the RNeasy Plant Mini Kit. The quality of RNA samples was assessed on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

Microarray analysis

The 44K Agilent microarray slides used in the present study were prepared in accordance with a previously reported method (Ishiguro et al., 2014). The labeling reactions were performed using a Quick Amp Labeling kit (Agilent Technologies). The Cy3-labeled cRNAs were fragmented and hybridized to the microarray slides at 65°C for 17 h. One-color spike mix was added to total RNA samples before reactions and used to monitor the efficiency of the labeling reactions and hybridizations. Hybridization and washing conditions of the hybridized slides followed manufacturer's protocol. The slides were scanned on an G2565BA microarray scanner (Agilent). The background correction of raw Cy3 signals was conducted using FEATURE EXTRACTION, version 9.5.1 (Agilent). All microarray analyses were performed with three independent biological replicates of embryo and endosperm tissues.

Microarray data analysis

Data analysis was performed in R, version 3.5.2 (R Foundation, Vienna, Austria). To determine the difference in TE expression among 5, 7, and 10 DAP embryos and endosperm, the raw signal data were normalized using the quantile method in the R package *limma*, version 3.38.3 (Ritchie et al., 2015). The individual $\log_2$-transformed normalized signal densities that were the mean of signals of negative-control probes were used as expression levels. Differentially expressed (DE) probes between the embryo and endosperm at 5, 7, and 10 DAP were identified with the row *t*-test in the R package *genefilter*, version 1.64.0 (Gentleman et al., 2021) and with the Storey Tibshirani method in the R package *qvalue*, version 2.14.1 (Storey et al., 2020). The false discovery rate (FDR) was set to 0.001. To analyze the differences in TE expression during endosperm development, raw data from 2 to 7 DAP and 10 DAP samples were normalized using the quantile method. DE probes during endosperm development were identified by one-way ANOVA and Storey methods with a FDR of 0.001. The number of *k*-means clusters of DE TEs, retrotransposons, DNA transposons, and MITEs was determined based on the *k*-gap score calculated by clusGap in the R package *cluster*, version 2.1.0 (Maechler et al., 2021; Yao, 2013). For comparison of the levels of *Youren* transcripts across rice tissues, the signal densities from leaves and anthers were retrieved from the microarray data series GSE49561 from the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo) at NCBI. The alignment of DE *Youren* probes was performed in CLUSTALW (http://www.genome.jp/tools-bin/clustalw), and the resulting phylogenetic tree was drawn with FIGTREE (http://tree.bio.ed.ac.uk). The genomic distribution of MITEs was plotted with the R packages *karyoploteR*, version 1.16.0 (Gel & Serra, 2017) and *BSgenome.Osativa.MSU.MSU7*, version 0.99.1 (Yao, 2013).

Statistical analysis

A Wilcoxon rank-sum test was used to assess the significance of differences in the distributions between different gene sets in the microarray analysis. The significance of the enrichment of a

specific DE TE subclass or MITE subfamily was tested using a hypergeometric test by calculating how frequently specific DE probes were enriched in all DE probes or all annotated repeats. The resulting *P*-values were adjusted using the Benjamini–Hochberg method. All statistical analyses were performed using R software.

RT-PCR analysis of *Youren*

First-strand cDNA was synthesized with the PrimeScript First Strand cDNA Synthesis Kit (Takara) with oligo-dT primers and total RNA isolated from 5 DAP endosperm. Oligo-dT was chosen here to match the conditions used for microarray labeling. First-strand cDNAs were used as template for PCR with the forward primer F1 (5′-CACCGGCAAGTGAATAAATGAGGAAAT-3′) and the reverse primer R (5′-TCCCAACCCAWRACACTAGACAT-3′), which were designed to anneal to the *Youren* copy with the highest transcript level in 5 DAP endosperm. PCR reactions were performed with SeqAmp DNA Polymerase (Takara), a proofreading polymerase. The 3′ end sequence of the F1 primer was set to one of the single nucleotide polymorphism sites of the *Youren* family to specifically amplify the most highly expressed copy (three copies in the rice genome). For the other *Youren* copies, the F2 primer (5′-CACCGGCAAGTGAATAAATGAGGAAAG-3′) was used instead, with PCR reactions using FastStart Taq DNA Polymerase (Sigma-Aldrich, St Louis, MO, USA), which does not have proofreading activity. Therefore, the single nucleotide polymorphism site cannot be corrected during the amplification, allowing the amplification of other copies expressed at lower levels. *ACTIN* was used as a control for RT-PCR (Ishikawa et al., 2011). *Youren* RT-PCR products were cloned into the pENTR™/D-TOPO™ vector (Takara) and sequenced by Sanger sequencing (model 3500; Applied Biosystems).

In situ hybridization

In situ hybridization was performed as described previously (Ishimoto et al., 2019). The sense and antisense probes were synthesized from a cDNA clone of *Youren* (ORSiTEMT0520001), which had been amplified with the forward primer F2 and the reverse primer R, as described above. The probes were hybridized to transverse sections of seed tissue at 4 DAP.

ACKNOWLEDGEMENTS

We thank Dr Atsushi Hoshino and Dr Takuji Tsukiyama for their valuable comments on our manuscript. This work was partly supported by a Grand-in-Aid for Scientific Research on Innovative Area (16H06464, 16H06471 and 16H21727 to TKi, 19H04873 to TKa), Grant-in-Aid for Scientific Research (C) (19 K05974 to AO) and a JSPS Research Fellowship (16 J02580 to KT) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

ACCESSION NUMBERS

The microarray data obtained in the present study have been deposited in the NCBI Gene Expression Omnibus (GEO) database under accession number GSE185870.

AUTHOR CONTRIBUTIONS

HN, AO, KT, and TKi conceived the research. AO, KT, TKa, and KY supervised the experiments. HN and YS performed experiments. YK provided materials. HN drafted the manuscript, designed the experiments, and analyzed the data. AO and TKi helped design the experiments and revise the draft. TKi wrote the article with contributions from all authors.

CONFLICT OF INTERESTS

The authors declare no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Expression analysis of genes and TEs during endosperm development.

Figure S2. Expression levels of *Youren* elements are high in the endosperm.

Figure S3. RT-PCR analysis of *Youren* transcripts.

Figure S4. *In situ* hybridization of *Youren* in the endosperm.

Figure S5. Temporal expression patterns and transcriptional direction of MITE subfamilies.

Figure S6. Distribution of *Youren*, *Tourist*, *Castaway*, *MITE-adh type M*, *Gaijin*, and *Ditto* copies corresponding to the microarray probes.

Figure S7. Proportion of *Youren* orientation relative to the nearest transcriptional unit.

Figure S8. Transcripts of the *NF-YA* genes in the embryo and endosperm.

Table S1. List of DEGs and DE TEs.

Table S2. Enrichment analysis of DE probes and individual TE classes.

Table S3. List of DE TEs and repeats in the *k*-means clusters.

Table S4. List of DE MITEs in the *k*-means clusters.

OPEN RESEARCH BADGES

This article has earned an Open Data and Open Materials badges. Data and materials are available at NCBI Gene Expression Omnibus (GEO), under accession number GSE185870 (https://www.ncbi.nlm.nih.gov/geo/).

DATA AVAILABILITY STATEMENT

All relevant data can be found within the published article and its supporting material, and GEO database. The experimental material should be requested to the corresponding author.

REFERENCES

Arabidopsis Genome Initiative. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Bai, A.N., Lu, X.D., Li, D.Q., Liu, J.X. & Liu, C.M. (2016) NF-YB1-regulated expression of sucrose transporters in aleurone facilitates sugar loading to rice endosperm. *Cell Research*, **26**, 384–388.

Bartels, A., Han, Q., Nair, P., Stacey, L., Gaynier, H., Mosley, M. *et al.* (2018) Dynamic DNA methylation in plant growth and development. *International Journal of Molecular Sciences*, **19**, 2144.

Batista, R.A. & Köhler, C. (2020) Genomic imprinting in plants-revisiting existing models. *Genes & Development*, **34**, 24–36.

Berger, F., Hamamura, Y., Ingouff, M. & Higashiyama, T. (2008) Double fertilization - caught in the act. *Trends in Plant Science*, **13**, 437–443.

Boulikas, T. (1995) Chromatin domains and prediction of MAR sequences. *International Review of Cytology*, **162A**, 279–388.

Bourc'his, D. & Voinnet, O. (2010) A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science*, **330**, 617–622.

Cho, J. & Paszkowski, J. (2017) Regulation of rice root development by a retrotransposon acting as a microRNA sponge. *eLife*, **6**, e30038.

Doring, H.P. & Starlinger, P. (1986) Molecular genetics of transposable elements in plants. *Annual Review of Genetics*, **20**, 175–200.

Feng, Y. (2003) Plant MITEs: useful tools for plant genetics and genomics. *Genomics, Proteomics & Bioinformatics*, **1**, 90–100.

Feschotte, C. (2008) Transposable elements and the evolution of regulatory networks. Nature Reviews. *Genetics*, **9**, 397–405.

Feschotte, C., Jiang, N. & Wessler, S.R. (2002) Plant transposable elements: where genetics meets genomics. *Nature Reviews. Genetics*, **3**, 329–341.

Feschotte, C. & Wessler, S.R. (2002) Mariner-like transposases are widespread and diverse in flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 280–285.

Frost, J.M., Kim, M.Y., Park, G.T., Hsieh, P.H., Nakamura, M., Lin, S.J.H. *et al.* (2018) FACT complex is required for DNA demethylation at heterochromatin during reproduction in Arabidopsis. *Proceedings of the National Academy of Sciences of the United States of America*, **115**, E4720–E4729.

Fujimoto, R., Sasaki, T., Kudoh, H., Taylor, J.M., Kakutani, T. & Dennis, E.S. (2011) Epigenetic variation in the FWA gene within the genus Arabidopsis. *The Plant Journal*, **66**, 831–843.

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

Gel, B. & Serra, E. (2017) karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics*, **33**, 3088–3090.

Gentleman, R., Carey, V., Huber, W. and Hahne, F. (2021) Genefilter: genefilter: methods for filtering genes from high-throughput experiments. R package version 1.72.1.

Gierl, A., Lutticke, S. & Saedler, H. (1988) TnpA product encoded by the transposable element En-1 of *Zea mays* is a DNA binding protein. *The EMBO Journal*, **7**, 4045–4053.

Hamant, O. & Pautot, V. (2010) Plant development: a TALE story. *Comptes Rendus Biologies*, **333**, 371–381.

Hashida, S.N., Uchiyama, T., Martin, C., Kishima, Y., Sano, Y. & Mikami, T. (2006) The temperature-dependent change in methylation of the antirrhinum transposon Tam3 is controlled by the activity of its transposase. *Plant Cell*, **18**, 104–118.

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

Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J. *et al.* (2011) Regulation of imprinted gene expression in Arabidopsis endosperm. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 1755–1762.

Ibarra, C.A., Feng, X., Schoft, V.K., Hsieh, T.F., Uzawa, R., Rodrigues, J.A. *et al.* (2012) Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science*, **337**, 1360–1364.

Ikeda, Y., Kinoshita, Y., Susaki, D., Ikeda, Y., Iwano, M., Takayama, S. *et al.* (2011) HMG domain containing SSRP1 is required for DNA demethylation and genomic imprinting in Arabidopsis. *Developmental Cell*, **21**, 589–596.

International Rice Genome Sequencing. (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

Ishiguro, S., Ogasawara, K., Fujino, K., Sato, Y. & Kishima, Y. (2014) Low temperature-responsive changes in the anther transcriptome's repeat sequences are indicative of stress sensitivity and pollen sterility in rice strains. *Plant Physiology*, **164**, 671–682.

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

Ishimoto, K., Sohonahra, S., Kishi-Kaboshi, M., Itoh, J.I., Hibara, K.I., Sato, Y. *et al.* (2019) Specification of basal region identity after asymmetric zygotic division requires mitogen-activated protein kinase 6 in rice. *Development*, **146**, dev176305.

Itoh, J., Sato, Y., Sato, Y., Hibara, K., Shimizu-Sato, S., Kobayashi, H. *et al.* (2016) Genome-wide analysis of spatiotemporal gene expression patterns during early embryogenesis in rice. *Development*, **143**, 1217–1227.

Jiang, N., Feschotte, C., Zhang, X. & Wessler, S.R. (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Current Opinion in Plant Biology*, **7**, 115–119.

Kawakatsu, T., Nery, J.R., Castanon, R. & Ecker, J.R. (2017) Dynamic DNA methylation reconfiguration during seed development and germination. *Genome Biology*, **18**, 171.

Kyozuka, J., Olive, M., Peacock, W.J., Dennis, E.S. & Shimamoto, K. (1994) Promoter elements required for developmental expression of the maize Adh1 gene in transgenic rice. *Plant Cell*, **6**, 799–810.

Lauria, M., Rupe, M., Guo, M., Kranz, E., Pirona, R., Viotti, A. *et al.* (2004) Extensive maternal DNA hypomethylation in the endosperm of *Zea mays*. *Plant Cell*, **16**, 510–522.

Lopes, M.A. & Larkins, B.A. (1993) Endosperm origin, development, and function. *Plant Cell*, **5**, 1383–1399.

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M. and Hornik, K. (2021) Cluster: cluster analysis basics and extensions. R package version 2.1.1.

Mantovani, R. (1999) The molecular biology of the CCAAT-binding factor NF-Y. *Gene*, **239**, 15–27.

Mirouze, M. & Vitte, C. (2014) Transposable elements, a treasure trove to decipher epigenetic variation: insights from Arabidopsis and crop epigenomes. *Journal of Experimental Botany*, **65**, 2801–2812.

Naito, K., Zhang, F., Tsukiyama, T., Saito, H., Hancock, C.N., Richardson, A.O. *et al.* (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*, **461**, 1130–1134.

Nardini, M., Gnesutta, N., Donati, G., Gatta, R., Forni, C., Fossati, A. *et al.* (2013) Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. *Cell*, **152**, 132–143.

Niu, B., Deng, H., Li, T., Sharma, S., Yun, Q., Li, Q. *et al.* (2020) OsbZIP76 interacts with OsNF-YBs and regulates endosperm cellularization in rice (*Oryza sativa*). *Journal of Integrative Plant Biology*, **62**, 1983–1996.

Niu, B., Zhang, Z., Zhang, J., Zhou, Y. & Chen, C. (2021) The rice LEC1-like transcription factor OsNF-YB9 interacts with SPK, an endosperm-specific sucrose synthase protein kinase, and functions in seed development. *The Plant Journal*, **106**, 1233–1246.

Ohnishi, T., Yoshino, M., Yamakawa, H. & Kinoshita, T. (2011) The biotron breeding system: a rapid and reliable procedure for genetic studies and breeding in rice. *Plant & Cell Physiology*, **52**, 1249–1257.

Oki, N., Yano, K., Okumoto, Y., Tsukiyama, T., Teraishi, M., Tanisaka, T.J.G. *et al.* (2008) A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. *japonica*. *Genes & Genetic Systems*, **83**, 321–329.

Ono, A. & Kinoshita, T. (2021) Epigenetics and plant reproduction: multiple steps for responsibly handling succession. *Current Opinion in Plant Biology*, **61**, 102032.

Oshlack, A. & Wakefield, M.J. (2009) Transcript length bias in RNA-seq data confounds systems biology. *Biology Direct*, **4**, 14.

Ouyang, S. & Buell, C.R. (2004) The TIGR plant repeat databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Research*, **32**, D360–D363.

Park, K., Kim, M.Y., Vickers, M., Park, J.S., Hyun, Y., Okamoto, T. *et al.* (2016) DNA demethylation is initiated in the central cells of Arabidopsis and rice. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 15138–15143.

Qiu, Y. & Kohler, C. (2020) Mobility connects: transposable elements wire new transcriptional networks by transferring transcription factor binding motifs. *Biochemical Society Transactions*, **48**, 1005–1017.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W. *et al.* (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, **43**, e47.

Rodrigues, J.A., Ruan, R., Nishimura, T., Sharma, M.K., Sharma, R., Ronald, P.C. *et al.* (2013) Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 7934–7939.

Sabelli, P.A. & Larkins, B.A. (2009) The development of endosperm in grasses. *Plant Physiology*, **149**, 14–26.

Song, X. & Cao, X. (2017) Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. *Current Opinion in Plant Biology*, **36**, 111–118.

Storey, J.D., Bass, J.A., Dabney, A. and Robinson, D. (2020) Qvalue: Q-value estimation for false discovery rate control. R package version 2.22.0. http://github.com/jdstorey/qvalue

Tao, Z., Shen, L., Gu, X., Wang, Y., Yu, H. & He, Y. (2017) Embryonic epigenetic reprogramming by a pioneer transcription factor in plants. *Nature*, **551**, 124–128.

Tonosaki, K., Ono, A., Kunisada, M., Nishino, M., Nagata, H., Sakamoto, S. *et al.* (2021) Mutation of the imprinted gene OsEMF2a induces autonomous endosperm development and delayed cellularization in rice. *Plant Cell*, **33**, 85–103.

Wells, J.N. & Feschotte, C. (2020) A field guide to eukaryotic transposable elements. *Annual Review of Genetics*, **54**, 539–561.

Xing, M.Q., Zhang, Y.J., Zhou, S.R., Hu, W.Y., Wu, X.T., Ye, Y.J. *et al.* (2015) Global analysis reveals the crucial roles of DNA methylation during rice seed development. *Plant Physiology*, **168**, 1417–1432.

Xu, Y., Gan, E.S., He, Y. & Ito, T. (2013) Flowering and genome integrity control by a nuclear matrix protein in Arabidopsis. *Nucleus*, **4**, 274–276.

Yang, G., Lee, Y.H., Jiang, Y., Shi, X., Kertbundit, S. & Hall, T.C. (2005) A two-edged role for the transposable element kiddo in the rice ubiquitin2 promoter. *The Plant Cell*, **17**, 1559–1568.

Yao, W. (2013) BSgenome.Osativa.MSU.MSU7: Oryza sativa full genome (MSU7). R package version 0.99.1.

Yuan, J., Chen, S., Jiao, W., Wang, L., Wang, L., Ye, W. *et al.* (2017) Both maternally and paternally imprinted genes regulate seed development in rice. *The New Phytologist*, **216**, 373–387.

Zemach, A., Kim, M.Y., Silva, P., Rodrigues, J.A., Dotson, B., Brooks, M.D. *et al.* (2010) Local DNA hypomethylation activates genes in rice endosperm. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 18729–18734.

Zhang, H., Lang, Z. & Zhu, J.K. (2018) Dynamics and function of DNA methylation in plants. *Nature Reviews. Molecular Cell Biology*, **19**, 489–506.

Zhang, X., Jiang, N., Feschotte, C. & Wessler, S.R. (2004) PIF-and pong-like transposable elements: distribution, evolution and relationship with tourist-like miniature inverted-repeat transposable elements. *Genetics*, **166**, 971–986.

Zhigou, E., Li, T., Zhang, H., Liu, Z., Deng, H., Sharma, S. *et al.* (2018) A group of nuclear factor Y transcription factors are sub-functionalized during endosperm development in monocots. *Journal of Experimental Botany*, **69**, 2495–2510.