Title:

Temporal Changes in Transcripts of MITE Transposable Elements during Rice Endosperm Development

Short title:

Characterization of MITE transcripts in Rice Endosperm

Authors:

Hiroki Nagata¹, Akemi Ono¹, Kaoru Tonosaki^{1,2}, Taiji Kawakatsu³, Kentaro Yano⁴, Yuji Kishima⁵, Tetsu Kinoshita^{1,6}

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 Department of Life Sciences, School of Agriculture, Meiji University, 1-1-1 Higashi-mita, Kawasaki 214-8571, Japan.

⁵ Research Faculty of Agriculture, Hokkaido University, Kita-9 Nishi-9, Kita-ku, Sapporo 060-8589, Japan

⁶ Corresponding Authors: tkinoshi@yokohama-cu.ac.jp

Author contributions:

H.N.; A.O.; K.T.; T.Ki. conceived the research; A.O.; K.T.; T.Ka.; K.Y. supervised the experiments; H.N. performed all experiments; Y.K. provided materials.; H. N. drafted the manuscript, designed the experiments and analyzed the data; A.O.; T.Ki. helped design experiments and revise the draft; T.Ki. wrote the article with contributions from all authors.

31 **ABSTRACT**

32 The repression of transcription from transposable elements (TEs) by DNA methylation is
33 necessary to maintain genome integrity and prevent harmful mutations. However, under
34 certain circumstances, TEs are thought to escape from the host defense system and
35 reactivate their transcription. In *Arabidopsis* (*Arabidopsis thaliana*) and rice (*Oryza*
36 *sativa*), DNA demethylases target the sequences derived from TEs in the central cell, the
37 progenitor cell for the endosperm in the female gametophyte. This genome-wide DNA
38 demethylation is also observed in the endosperm after fertilization. In this study, we used
39 a custom microarray to survey the transcripts generated from TEs during the rice
40 endosperm development and at selected timepoints in the embryo as a control. The
41 expression patterns of TE transcripts are dynamically up- and downregulated during
42 endosperm development, especially for miniature inverted-repeat transposable elements
43 (MITEs). Surprisingly, some TE transcripts were directionally controlled, while the
44 other DNA transposons and retrotransposons were not. We also discovered the NF-Y
45 binding motif, CCAAT, in the region near the 5' terminal inverted repeat of *Youren*, one
46 of the transcribed MITEs in the endosperm. Our results uncover dynamic changes in TE
47 activity during endosperm development in rice.

48 INTRODUCTION

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

60 The endosperm is the nutritive tissue that sustains embryonic growth during seed
61 development and seedling growth after germination, especially in crop species such as rice
62 (*Oryza sativa*) (Lopes and Larkins, 1993; Sabelli and Larkins, 2009). The endosperm arises
63 from the second fertilization event between the central cell of the female gametophyte and one
64 of the two sperm cells delivered by the male pollen tube (Berger et al., 2008). The embryo is
65 itself the product of the fertilization of the egg and the other sperm cell. Unlike the embryo, the
66 endosperm has unique characteristics; its genome comprises two maternal copies and one
67 paternal copy, undergoes terminal differentiation, and is subjected to global DNA demethylation
68 (Lauria et al., 2004; Hsieh et al., 2009; Zemach et al., 2010).

69 In *Arabidopsis* (*Arabidopsis thaliana*), one of the DNA glycosylase termed DEMETER
70 (DME) targets AT-rich transposons in euchromatin (Gehring et al., 2006; Ibarra et al.,
71 2012) as well as the heterochromatic TEs with the histone chaperone FACT in the central
72 cell, the progenitor cell of the endosperm before fertilization (Ikeda et al., 2011; Frost et
73 al., 2018). This active demethylation also contributes to parent-of-origin specific gene
74 expression, known as imprinting, in the endosperm (Fujimoto et al., 2011; Hsieh et al.,

2011), possibly via an association with TEs (Qiu and Kohler, 2020; Rodrigues et al., 2021). Indeed, many TEs are thought to be activated in the endosperm, which is consistent with pioneering work exploring the causes of kernel pigmentation patterns (Doring and Starlinger, 1986) and maternal DNA hypomethylation in maize (*Zea mays*) endosperm (Lauria et al., 2004). While 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 that 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 and Voinnet, 2010). In addition, the endosperm is a terminally differentiated tissue, which might therefore tolerate high TE activity brought upon by DNA demethylation. De-repression of TE activity influences the neighboring genomic regions (Yang et al., 2005; Naito et al., 2009), 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 about three times larger than that of Arabidopsis (125 Mb) (Arabidopsis Genome Initiative, 2000; International Rice Genome Sequencing, 2005). However, rice contains much more repetitive sequences than does Arabidopsis (Song and Cao, 2017). The TE distribution is also different: TEs in Arabidopsis are mainly located in the pericentromeric regions, while those in rice also populate in the euchromatin (Mirouze and 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 transposable elements (MITEs) (Jiang et al., 2004). MITEs are internal deletion derivatives of autonomous DNA transposons with terminal inverted repeats (TIRs) at both ends (Feschotte et al., 2002; Feschotte and Wessler, 2002; Feng, 2003). These short stretches of sequences, 100–600 bp in length,

101 mostly reside in the euchromatic regions and are relatively close to genes compared to longer
102 long terminal repeat (LTR)-type retrotransposons (Oki et al., 2008; Zemach et al., 2010).

103 In this study, we performed a time-course transcriptome analysis of rice TEs during
104 endosperm development. We employed a custom-made microarray designed to detect
105 transcripts generated from repetitive sequences in rice (Ishiguro et al., 2014), to bypass
106 the limitations associated with short-read sequencing of TEs, especially for short and
107 high-copy number elements like MITEs. We determined that many expressed TEs in the
108 endosperm tended not to be expressed in the embryo and showed a temporal dependent
109 expression pattern. Among expressed TEs, many MITEs displayed a directional behavior,
110 as we detected transcripts originating from one strand only, likely reflecting the direction
111 of the ancestral autonomous transcriptional unit. This pattern was not present in other
112 tested DNA transposons and retrotransposons. We also identified the binding motif for
113 NUCLEAR FACTOR Y (NF-Y) transcription factors (Mantovani, 1999; Nardini et al.,
114 2013) within *Youren* sequences, which overlapped with the boundary of the 5' TIR and
115 the internal region. Together, our results suggest that while MITEs are inserted proximal
116 to genes or coding regions in random orientations, MITE transcripts do not hitchhike the
117 directional transcription of their neighboring genes (i.e., readthrough or chimeric
118 transcripts) but rather exhibit their own directional transcriptional control, possibly
119 starting from intrinsic sequences in *Youren* MITEs. Such an active transcriptional
120 environment at MITE sequences may contribute to shaping the transcriptional network
121 of genes expressed in the rice endosperm.

122

123

124 RESULTS

125 Expression of TEs in the embryo and the endosperm

126 Genome-wide hypomethylation and TE transposition are prominent features of the
 127 endosperm, but how TE expression changes over time during seed development is not
 128 well described. Therefore, we conducted a transcriptome analysis with a custom 44k
 129 Agilent microarray (Ishiguro et al., 2014) harboring 31,366 repeat-related, 864
 130 microRNA (miRNA)-related probes, as well as probes for 9,563 genes. To investigate
 131 the dynamics of TE transcription, we isolated total RNAs from rice endosperm in 24-h
 132 intervals from 2 to 7 days after pollination (DAP) and at 10 DAP (Supplemental Figure
 133 S1A). The endosperm became coenocytic at 2 DAP and grew in a centripetal direction by
 134 3 DAP. Cellularization took place between these time points in our growth conditions
 135 (Tonosaki et al., 2021). The endosperm continued to divide and began to deposit storage
 136 compounds until the maturation stage at 10 DAP (Supplemental Figure S1A). Our chosen
 137 time points thus covered most developmental transitions in endosperm development. In
 138 addition to the endosperm, we also extracted total RNAs from developing embryos
 139 collected at 5, 7 and 10 DAP as controls, which correspond to the coleoptile, the
 140 completion of organogenesis, and the maturation stages of rice embryogenesis,
 141 respectively (Itoh et al., 2016) (Supplemental Figure S1B).

142 We performed a principal component analysis (PCA) using the expression of genes
 143 and TEs obtained from all samples across replicates. We observed separation of genes
 144 and TEs expressed in the embryo from those expressed in the endosperm, as well as high
 145 reproducibility across biological replicates (Supplemental Figure S1C). In addition,
 146 marker genes for embryo or endosperm exhibited their expected expression patterns
 147 (Supplemental Figure S1D). Using matched sample pairs from embryo and endosperm
 148 collected at 5, 7 and 10 DAP, we drew MA plots to explore the underlying expression
 149 differences for TEs. While we detected 105 (5 DAP) and 554 (7 DAP) differentially
 150 expressed (DE) TEs during earlier stages of embryo and endosperm development, we

151 identified vastly more DE TEs ($n = 2,123$) in the latest stage (Figure 1, A–C). We
 152 subdivided probes into “upregulated in embryo” and “upregulated in endosperm”
 153 categories for each tissue (Figure 1, D–F), as TEs may be expressed at low levels in the
 154 embryo if they are highly expressed in the endosperm, according to the proposed
 155 hypothesis. Although we observed many upregulated TEs in the embryo in our matched
 156 sample pairs at 5, 7 and 10 DAP, many upregulated TEs in the endosperm showed a
 157 correspondingly low expression level in the embryo (Figure 1, D–F). These results were
 158 consistent with the proposed hypothesis that TEs are highly transcribed to silence their
 159 counterparts in the embryo genome, although it is also possible that TE expression
 160 patterns merely reflect the differences in the developmental program of the embryo and
 161 endosperm. In any case, we observed that the expression patterns of TEs were largely
 162 distinct between the embryo and the endosperm, as also seen for protein-coding genes
 163 (Figure 1, A–C; Supplemental Figure S1D).

164

165 **Temporal Expression Patterns of TEs during Rice Endosperm Development**

166 We focused on all samples derived from the endosperm to investigate the temporal
 167 expression patterns of the repetitive sequences, including those of TEs, during rice
 168 endosperm development from 2 to 10 DAP (Figure 2, A). We determined differentially
 169 expressed probes including TEs and repeats during endosperm development by one-way
 170 ANOVA and Storey methods (see Materials and Methods) in this data set. We detected
 171 dynamic and stage-specific expression patterns when taking a k-means clustering
 172 approach, which produced 10 distinct clusters. Clusters A3, A6 and A7 (Figure 2, A)
 173 showed an upregulation at 2 and 3 DAP, which may reflect DNA demethylation of the
 174 central cell and early endosperm development. However, the majority of repeats
 175 appeared to be upregulated at later stages between 7 and 10 DAP, as illustrated by clusters
 176 A2, A4, A8, A9 and A10 (Figure 2, A). This result was consistent with the pairwise
 177 comparisons of the embryo and endosperm samples at 10 DAP, at which stage more TEs

178 were activated in the endosperm (Figure 1, C and F). Among the DE TEs during
 179 endosperm development, the MITE family was prominent in each cluster rather than
 180 retrotransposons, DNA transposons, or other repeats (Figure 2, B). We thus extracted the
 181 data from MITE probes only for reanalysis (Figure 3, A) and sorted it individual MITE
 182 subfamilies (Figure 3, B). As a result, we assessed detectable expression for the MITE
 183 subfamilies *Tourist*, *Castaway*, *MITE-adh* Type M, *Gaijin/Gaigin* and *Ditto*. These
 184 MITE subfamilies were differentially expressed during endosperm development, with
 185 broad expression patterns, as all clusters contained representative members of all
 186 subfamilies. By contrast, a set of probes corresponding to *Youren* MITEs were
 187 predominantly enriched in the M6 cluster (Figure 3, C). *Youren* belongs to the
 188 *Tourist*-like elements, with 3-bp (TAA or TTA) target site duplications (Zhang et al.,
 189 2004). These elements were characterized by one of the highest signal intensities among
 190 all DE TEs in the endosperm at 5 DAP (Figure 3, D) and fairly low signal intensities in
 191 the embryo, anthers and leaves (Supplemental Figure S2).

192 To assess the possibility of cross-hybridization of highly conserved repetitive
 193 transcripts like those derived from TEs, we investigated the relationship between the
 194 expression levels and the level of identity between probes (Figure 4, A and B). This
 195 possibility was of special concern for *Youren* elements, as they showed overlapping
 196 expression patterns but were limited to a single cluster (Figure 3, A and B). We therefore
 197 subdivided all *Youren* elements into 11 groups based on their expression level (Figure 4,
 198 A) and generated a phylogenetic tree of each probe recognizing each *Youren* element,
 199 which revealed no correlation between sequence similarity and signal intensity except
 200 for the 10th and 11th weakly expressed groups (Figure 4, B). Therefore, although we do
 201 not completely exclude the possibility of cross-hybridization for a small fraction of
 202 probes, we conclude that each probe detects transcripts generated from multiple loci in
 203 the rice genome.

204 Since MITE sequences are deposited in genome databases without regard for their
 205 orientation (Ouyang and Buell, 2004), reference MITE sequences comprise a mixture of
 206 forward and reverse orientations (Zhang et al., 2004). Indeed, a phylogenetic tree built
 207 from full-length *Youren* sequences in databases formed two distinct clades (Figure 4, C).
 208 To determine the direction of *Youren* transcripts, we reannotated each *Youren* probe as
 209 reference and reverse orientation, according to the orientation of each *Youren* element
 210 deduced from the deposited sequence, taking the direction of transcripts from the
 211 putative autonomous copy as a guide. The 60-mer probes printed onto the microarrays
 212 were designed to hybridize to the corresponding 5' and 3' end sequences of annotated
 213 *Youren* elements, making it possible for probes to detect antisense and sense transcripts
 214 (Ishiguro et al., 2014) (Figure 4, D). Surprisingly, most *Youren* transcripts appeared to
 215 hybridize to the 3' probes, suggesting that their transcripts are in the reference
 216 orientation (Figure 4, D and E). We independently confirmed that *Youren* transcripts are
 217 transcribed in the reference strand by RT-PCR (Supplemental Figure S3). Thus, our
 218 results suggest the coordinated expression of *Youren* elements originating from different
 219 copies scattered over the rice genome. We observed a similar trend, that is, temporal
 220 expression patterns were associated with the direction of transcription, for *Tourist*,
 221 *Castaway*, *MITE-adh* (type M), and *Gaijin* MITE elements, but not for *Ditto*
 222 (Supplemental Figure S4).

223

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

225 We sought to understand the basis for the directional transcription of MITEs.
 226 Although we could not distinguish individual insertions of multicopy MITE sequences
 227 from the microarray data, we determined locations of all copies of DE TEs and non-DE
 228 TEs in the rice genome. The resulting karyoplots indicated that all tested MITE
 229 subfamilies map to all chromosomal arm regions, in agreement with a previous report
 230 (Mirouze and Vitte, 2014), for both DE MITEs and non-DE MITEs (Supplemental Figure

231 S5). We again focused on highly expressed *Youren* (note that we used 198 different DE
232 *Youren* copies, since eight copies were detected by probes hybridizing to both ends,
233 pointing to these copies as duplications). We also failed to detect any significant skew
234 between MITE subfamilies and DE and non-DE *Youren* elements, for the position of their
235 insertion sites relative to the upstream, coding, intron or downstream region of
236 protein-coding genes (Figure 5, A). The orientations of *Youren* insertions relative to the
237 direction of transcription for their closest gene appeared to be random (Supplemental
238 Figure S6). We thus concluded that these directional transcripts likely arose from
239 internal *Youren* sequences.

240 Since transcript directionality can be attributed to the presence of *cis*-regulatory
241 elements in 5' sequences, we searched for such elements in all *Youren* copies. A CCAAT
242 motif, which is the typical binding site for NF-Y transcription factors, was present at the
243 boundary between the 5' TIR and the internal sequence of *Youren*. We also identified a
244 fragment of this motif in the 3' TIR as a consequence of the inverted repeat, although the
245 internal sequence did not provide the necessary sequence to reconstitute the entire motif
246 (Figure 5, B). We extended our *in silico* analysis to the entire *Youren* sequences using the
247 online tool Plant Promoter Analysis Navigator (PlantPAN 3.0;
248 <http://plantpan.itps.ncku.edu.tw>), which identified Homeodomain TALE,
249 GT1CONSENSUS, and AT-Hook *cis*-elements (Kyozyuka et al., 1994; Boulikas, 1995;
250 Hamant and Pautot, 2010; Xu et al., 2013). Surprisingly, these transcription factor
251 binding sites were preferentially located in the 5' region of *Youren* for forward
252 orientation, but not for reverse orientation of *Youren* (Figure 5, C) or other tested MITEs.
253 For *Youren* elements with detectable transcripts by RT-PCR, the AT-Hook binding sites
254 located within the forward primer used for PCR, suggesting that the element overlaps
255 with the transcribed *Youren* region (Supplemental Figure S3). The NF-Y binding motif is
256 a good candidate to explain the directional transcription of *Youren*, as several genes
257 encoding the trimeric NF-Y complex, *NF-YB1*, *NF-YB7*, and *NF-YB9*, are predominantly

258 expressed in the aleurone and starchy endosperm and control seed development in rice
 259 (Bai et al., 2016; Niu et al., 2020; Niu et al., 2021). In addition, the temporal expression
 260 patterns of *NF-YB1*, *NF-YB7*, and *NF-YB9* are similar to those of *Youren* elements. In
 261 conclusion, we identified coordinated expression patterns of *Yourens* and the potential
 262 underlying *cis*-elements.

263

264

265 **DISCUSSION**

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

278 Since genome-wide demethylation has been observed in the central cell and endosperm
 279 cells, a simple prediction is that TE activity would be high in the endosperm at all
 280 developmental stages (Ibarra et al., 2012; Park et al., 2016). However, our temporal
 281 analysis revealed a very different picture, as the expression of TEs was not static and
 282 high but exhibited dynamic changes, with some TEs expressed during earlier stages and
 283 others expressed later. We observed a similar trend in the embryo, although we limited
 284 our investigation to three time points, representing the coleoptile, the completion of
 285 organogenesis and the maturation stages of rice embryogenesis. Over the course of these
 286 time points, the embryo develops into a mature structure that consists of several tissues:
 287 the root and shoot primordium, leaves and the scutellum. The expression of rice TEs was
 288 reported to be largely different in distinct plant tissues (Cho and Paszkowski, 2017); our
 289 results therefore likely merely reflect the tissue- and organ-specific transcriptional
 290 programs in the embryo. Similarly, the rice endosperm develops into several domains,
 291 such as the peripheral region and the region adjacent to the embryo, as well as the

292 aleurone layer and the starchy endosperm, and the expression profile of TEs may differ
 293 in each compartment. Alternatively, the transcriptional environment of TEs may rely on
 294 temporal programs and epigenetic changes of TE sequences. Demethylation of the
 295 maternal genome has been observed in several plant species, resulting in imprinted gene
 296 expression (Hsieh et al., 2009; Rodrigues et al., 2013; Batista and Kohler, 2020). *Youren*
 297 has the potential to contribute to this type of epigenetic modulation of gene expression
 298 (Yuan et al., 2017). In addition to the demethylation of the maternal copy of the
 299 endosperm genome, DNA methylation status changes during seed and endosperm
 300 development in both Arabidopsis and rice (Xing et al., 2015; Kawakatsu et al., 2017). To
 301 better understand the roles of TEs and demethylation in gene expression during
 302 endosperm development, it is necessary to conduct higher-resolution expression
 303 analyses at the cell type level, combined with multiple omics analyses to uncover the
 304 epigenetic states underlying endosperm development.

305 Although the transcription of long and almost intact TEs like LTR retrotransposons is
 306 directionally controlled, the transcription of remnant TEs is thought to be influenced by
 307 the expression of neighboring genes and may even form chimeric transcripts bringing
 308 together gene sequences and TEs. MITEs are remnant TEs derived from longer
 309 autonomous copies and thus no longer encode the sequences required for their
 310 transposition. Nevertheless, transcripts derived from MITE loci, with the exception of
 311 *Ditto*, are directionally controlled. Our results suggest that these elements might have
 312 preserved the promoter activity required for autonomous transcription. Among the tested
 313 MITEs, *Youren* was unique, as the expression of most elements was relatively high,
 314 similar and directionally controlled. We hypothesize that these patterns are caused by
 315 *cis*-elements that reside in the 5' internal region of *Youren*, since the sequences outside of
 316 TE insertion sites are not conserved and cannot explain their coordinated expression.

317 The approach we followed in this study provides the framework to test our proposed
 318 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 good candidate. The NF-YB and NF-YC subunits have histone-fold domains to bind to DNA in a non-sequence-specific manner, while 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. It will be worth investigating whether and how much these elements can rewire transcriptional networks during endosperm in the context of the evolutionary history of rice.

332

333 MATERIALS AND METHODS

334 Plant Materials and Growth Conditions

335 Rice plants (*Oryza sativa* cv. Nipponbare) were grown in a Biotron chamber
336 (NC-350HC; Nihon Ikakikai) according to a previously described protocol (Ohnishi et
337 al., 2011) adjusted for short-day photoperiod and temperature cycles (11-h light at
338 30°C/13-h dark at 25°C).

339

340 RNA Extraction from Rice Endosperm

341 Rice endosperm was harvested in 24-h intervals from developing seeds from 2 to 7 d
342 after pollination (DAP) and at 10 DAP (Supplemental Figure S1A). For the isolation of
343 liquid coenocytic endosperm (2 DAP) and just after cellularization (3 DAP), a cut was
344 made on the surface of the seed coat using a scalpel under a binocular microscope, as
345 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 μ L of the extraction buffer from the PicoPure RNA isolation kit (Applied Biosystems). 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 μ L of RLC buffer (Plant RNeasy kit; Qiagen), 100 μ L of Fruit-mate (Takara) and 4.5 μ L of β -mercaptoethanol (Wako) (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 according to 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 multi-beads shocker (MB755U, Yasui Kikai) at 1,500 rpm for 2 s. The resulting homogenized endosperm was resuspended in 1.3 mL of Fruit-mate. The mixture was subdivided into two 650- μ L aliquots and centrifuged at 12,000g for 5 min at 4°C. To 250 μ L of each supernatant, 400 μ L of RLT buffer from the RNeasy Plant Mini Kit (Qiagen) was added together with 4 μ L of β -mercaptoethanol (Wako). For DNase treatment, RNase-Free DNase Set (Qiagen) was used according to the manufacturer's instructions.

364

365 **RNA Extraction from Rice Embryos**

Rice embryos were dissected from developing ovaries at 5, 7 and 10 DAP (Supplemental Figure S1B) under a binocular microscope. Since early-stage embryos are small and difficult to dissect under a binocular microscope, two protocols were tested to isolate RNA from embryos at 5 DAP: the RNeasy Plant Mini Kit and the PicoPure RNA isolation kit (Thermo Fischer Science). 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 multi-beads shocker at 1,500 rpm for 2 s; the

subsequent steps followed the protocol described above. For the PicoPure RNA isolation kit (Thermo Fischer Science), three embryos at 5 DAP were homogenized in 10 μ L 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 according to 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)

Microarray Analysis

The 44k Agilent microarray slides used in this study were prepared according to a previously published method (Ishiguro et al., 2014). The labeling reactions were performed using a Quick Amp Labeling kit (Agilent). 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 the Agilent protocol. The slides were scanned on an Agilent G2565BA microarray scanner (Agilent). The background correction of raw Cy3 signals was conducted using the Agilent Feature Extraction software (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). To determine the difference in TE expression between 5, 7 and 10 DAP embryos and endosperm, the raw signal data were normalized by 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 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 by the Quantile method. Differentially expressed 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 differentially expressed 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) (Yao, 2013; Maechler et al., 2021). 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 at NCBI. The alignment of DE *Youren* probes was performed in ClustalW and the resulting phylogenetic tree was drawn with FigTree. The genomic distribution of MITEs was plotted with the R packages *karyoploteR* (version 1.16.0) (Gel and Serra, 2017) and *BSgenome.Osativa.MSU.MSU7* (version 0.99.1)(Yao, 2013).

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 *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 (SNP) 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), which does not have proofreading activity. Therefore, the SNP site cannot not 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-TOPOTM vector (Takara) and sequenced by Sanger sequencing (Applied Biosystems 3500).

437

438 FIGURE LEGENDS

439 **Figure 1. Expression patterns of genes and transposable elements (TEs) in the rice embryo**
 440 **and endosperm. (A-C)**, MA plots showing the expression levels of genes and TEs at 5 DAP
 441 **(A)**, 7 DAP **(B)** and 10 DAP **(C)**. The light and dark brown dots represent differentially
 442 expressed (DE) TEs and genes upregulated in the endosperm. The light and dark green dots
 443 represent DE TEs and genes upregulated in the embryo. The black dots indicate non-significant
 444 probes of TEs and genes ($FDR < 0.001$). **(D-F)**, Rain cloud plots showing the expression
 445 levels of DE TEs. “Up in Emb” and “Up in End” represent negative and positive LogFC dots
 446 from MA plots data. The light blue and light brown dots indicate individual expression levels in
 447 the embryo and endosperm fractions. **, P -value = 0.001; *, P -value = 0.05 using a Wilcoxon
 448 rank sum test.

449

450 **Figure 2. Temporal expression patterns of repetitive sequences. (A)**, Heatmap representation
 451 of k-means clustering results of the transcription patterns of DE TEs and repeats during
 452 endosperm development. The colored lines to the right denote representative TE and repeat
 453 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.

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. **(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 Log₂-transformed normalized signal intensity for each MITE subfamily in 5 DAP endosperm. At least one probe for each individual subfamily was differentially expressed in the endosperm from 2 DAP to 10 DAP. The *Youren* subfamily shows one of the highest expression levels.

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 Log₂-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 tree diagram. **(C)**, Phylogenetic tree illustrating the sequence similarity of the entire *Youren* region. Clades 1 and 2 are composed of complementary sequences. **(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 Log₂-transformed normalized signal intensity. Cyan, *Youren* in reference orientation; magenta, *Youren* in reverse orientation.

481 **Figure 5. *Youren* insertions in the genome and transcription factor binding motifs. (A),**
 482 Proportion of insertion sites relative to genes for all MITEs, all *Youren* sequences, as well as DE
 483 and non-DE *Youren* sequences. **(B),** Sequence logos for the 5' (upper) and 3' reverse
 484 complementary (bottom) sequences that include the TIR and the inner regions, drawn by MEME
 485 suite. The horizontal line above the logos indicates the conserved CCAAT motif, known as the
 486 NF-Y binding site. The box indicates the last T of the CCAAT motif, which is not conserved in
 487 the 3' sequences of *Youren*. **(C),** Stacked histogram showing the positions of representative
 488 transcription factor binding sites and the top-strand (left) and bottom-strand (right) of DE *Youren*.
 489 *In silico* analysis of transcription factor binding sites of DE *Youren* sequences ($n = 198$) was
 490 performed with PlantPAN 3.0 (<http://plantpan.itps.ncku.edu.tw>).
 491

492 **Supplemental Data**

493 The following supplemental materials are available.

494
 495 **Supplemental Figure S1.** Expression analysis of genes and TEs during endosperm
 496 development.

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

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

499 **Supplemental Figure S4.** Temporal expression patterns and transcriptional direction of *MITE*
 500 subfamilies.

501 **Supplemental Figure S5.** Distribution of *Youren*, *Tourist*, *Castaway*, *MITE*-adh (type M),
 502 *Gaijin* and *Ditto* copies corresponding to the microarray probes.

503 **Supplemental Figure S6.** Proportion of *Youren* orientation relative to the nearest
 504 transcriptional unit.

505

506 **ACKNOWLEDGMENTS**

507 This work was partly supported by a Grand-in-Aid for Scientific Research on Innovative
 508 Area (16H06464, 16H06471 and 16H21727 to T.Ki., 19H04873 to T. Ka.) and a JSPS
 509 Research Fellowship (16J02580 to K.T.) from the Ministry of Education, Culture, Sports,
 510 Science and Technology (MEXT) of Japan.

511

512 **AUTHOR CONTRIBUTIONS**

513 K.T., A.O. and T.K. conceived research plans; T.K. and L.C. supervised the experiments;
 514 K.T., H.N., M.K. and T.K. performed the experiments; K.T analyzed the data; H.F., K.N.,
 515 Y.S. and M.E. provided research materials; K.T. and T.K. wrote the paper.

516

517

Figure 1. Expression patterns of genes and transposable elements (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 expressed (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 LogFC dots from MA plots data. The light blue and light brown dots indicate individual expression levels in the embryo and endosperm fractions. **, P -value = 0.001; *, P -value = 0.05 using a Wilcoxon rank sum test.

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.

Figure 3. Temporal expression patterns of MITE subfamilies during rice endosperm development.

(A), Heat map 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. (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 Log₂-transformed normalized signal intensity for each MITE subfamily in 5 DAP endosperm. At least one probe for each individual subfamily was differentially expressed in the endosperm from 2 to 10 DAP. The *Youren* subfamily shows one of the highest expression levels.

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 Log₂-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 tree diagram. (C), Phylogenetic tree illustrating the sequence similarity of the entire *Youren* region. Clades 1 and 2 are composed of complementary sequences. (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

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 by MEME suite. 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 *Youren*. (C), 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>).

Parsed Citations

- Arabidopsis Genome Initiative (2000)** Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796-815
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Bai AN, Lu XD, Li DQ, Liu JX, Liu CM (2016)** NF-YB1-regulated expression of sucrose transporters in aleurone facilitates sugar loading to rice endosperm. *Cell Res* 26: 384-388
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Bartels A, Han Q, Nair P, Stacey L, Gaynier H, Mosley M, Huang QQ, Pearson JK, Hsieh TF, An YC, Xiao W (2018)** Dynamic DNA Methylation in Plant Growth and Development. *Int J Mol Sci* 19
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Batista RA, Kohler C (2020)** Genomic imprinting in plants-revisiting existing models. *Genes Dev* 34: 24-36
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Berger F, Hamamura Y, Ingouff M, Higashiyama T (2008)** Double fertilization - caught in the act. *Trends Plant Sci* 13: 437-443
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Boulikas T (1995)** Chromatin domains and prediction of MAR sequences. *Int Rev Cytol* 162A: 279-388
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Bourc'his D, Voinnet O (2010)** A small-RNA perspective on gametogenesis, fertilization, and early zygotic development. *Science* 330: 617-622
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Cho J, Paszkowski J (2017)** Regulation of rice root development by a retrotransposon acting as a microRNA sponge. *Elife* 6
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Doring HP, Starlinger P (1986)** Molecular genetics of transposable elements in plants. *Annu Rev Genet* 20: 175-200
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Feng Y (2003)** Plant MITEs: useful tools for plant genetics and genomics. *Genomics, proteomics & bioinformatics* 1: 90-100
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Feschotte C (2008)** Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 9: 397-405
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Feschotte C, Jiang N, Wessler SR (2002)** Plant transposable elements: where genetics meets genomics. *Nat Rev Genet* 3: 329-341
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Feschotte C, Wessler SR (2002)** Mariner-like transposases are widespread and diverse in flowering plants. *Proc Natl Acad Sci U S A* 99: 280-285
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Frost JM, Kim MY, Park GT, Hsieh PH, Nakamura M, Lin SJH, Yoo H, Choi J, Ikeda Y, Kinoshita T, Choi Y, Zilberman D, Fischer RL (2018)** FACT complex is required for DNA demethylation at heterochromatin during reproduction in *Arabidopsis*. *Proc Natl Acad Sci U S A* 115: E4720-E4729
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Fujimoto R, Sasaki T, Kudoh H, Taylor JM, Kakutani T, Dennis ES (2011)** Epigenetic variation in the FWA gene within the genus *Arabidopsis*. *Plant J* 66: 831-843
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Gehring M, Huh JH, Hsieh TF, Penterman J, Choi Y, Harada JJ, Goldberg RB, Fischer RL (2006)** DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124: 495-506
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Gel B, Serra E (2017)** karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 33: 3088-3090
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Gentleman R, Carey V, Huber W, Hahne F (2021)** genefilter: genefilter: methods for filtering genes from high-throughput experiments. *R package version 1.72.1*.
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Gierl A, Lutticke S, Saedler H (1988)** TnpA product encoded by the transposable element En-1 of *Zea mays* is a DNA binding protein. *EMBO J* 7: 4045-4053
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)
- Hamant O, Pautot V (2010)** Plant development: a TALE story. *C R Biol* 333: 371-381
Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hashida SN, 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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hsieh TF, Ibarra CA, Silva P, Zemach A, Eshed-Williams L, Fischer RL, Zilberman D (2009) Genome-wide demethylation of Arabidopsis endosperm. Science 324: 1451-1454

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Hsieh TF, Shin J, Uzawa R, Silva P, Cohen S, Bauer MJ, Hashimoto M, Kirkbride RC, Harada JJ, Zilberman D, Fischer RL (2011) Regulation of imprinted gene expression in Arabidopsis endosperm. Proc Natl Acad Sci U S A 108: 1755-1762

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ibarra CA, Feng X, Schoft VK, Hsieh TF, Uzawa R, Rodrigues JA, Zemach A, Chumak N, Machlicova A, Nishimura T, Rojas D, Fischer RL, Tamaru H, Zilberman D (2012) Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. Science 337: 1360-1364

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ikeda Y, Kinoshita Y, Susaki D, Ikeda Y, Iwano M, Takayama S, Higashiyama T, Kakutani T, Kinoshita T (2011) HMG domain containing SSRP1 is required for DNA demethylation and genomic imprinting in Arabidopsis. Dev Cell 21: 589-596

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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 Physiol 164: 671-682

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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. Plant J 65: 798-806

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Itoh J, Sato Y, Sato Y, Hibara K, Shimizu-Sato S, Kobayashi H, Takehisa H, Sanguinet KA, Namiki N, Nagamura Y (2016) Genome-wide analysis of spatiotemporal gene expression patterns during early embryogenesis in rice. Development 143: 1217-1227

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Jiang N, Feschotte C, Zhang X, Wessler SR (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). Curr Opin Plant Biol 7: 115-119

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kawakatsu T, Nery JR, Castanon R, Ecker JR (2017) Dynamic DNA methylation reconfiguration during seed development and germination. Genome Biol 18: 171

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Kyozuka J, Olive M, Peacock WJ, Dennis ES, Shimamoto K (1994) Promoter elements required for developmental expression of the maize Adh1 gene in transgenic rice. Plant Cell 6: 799-810

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Lauria M, Rupe M, Guo M, Kranz E, Pirone R, Viotti A, Lund G (2004) Extensive maternal DNA hypomethylation in the endosperm of Zea mays. Plant Cell 16: 510-522

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Lopes MA, Larkins BA (1993) Endosperm origin, development, and function. Plant Cell 5: 1383-1399

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Maechler M, Rousseeuw P, Struyf A, Hubert M, Hornik K (2021) cluster: Cluster Analysis Basics and Extensions. R package version 2.1.1.

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Mirouze M, Vitte C (2014) Transposable elements, a treasure trove to decipher epigenetic variation: insights from Arabidopsis and crop epigenomes. J Exp Bot 65: 2801-2812

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461: 1130-1134

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Nardini M, Gnesutta N, Donati G, Gatta R, Forni C, Fossati A, Vonnrhein C, Moras D, Romier C, Bolognesi M, Mantovani R (2013) Sequence-specific transcription factor NF-Y displays histone-like DNA binding and H2B-like ubiquitination. Cell 152: 132-143

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Niu B, Deng H, Li T, Sharma S, Yun Q, Li Q, E Z, Chen C (2020) OsbZIP76 interacts with OsNF-YBs and regulates endosperm cellularization in rice (*Oryza sativa*). J Integr Plant Biol 62: 1983-1996

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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. Plant J 106: 1233-1246

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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 Physiol 52: 1249-1257

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Oki N, Yano K, Okumoto Y, Tsukiyama T, Teraishi M, Tanisaka TJG, systems g (2008) A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa* ssp. japonica. Genes Genet Syst. 83: 321-329

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ono A, Kinoshita T (2021) Epigenetics and plant reproduction: Multiple steps for responsibly handling succession. Curr Opin Plant Biol 61: 102032

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ouyang S, Buell CR (2004) The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. Nucleic Acids Res 32: D360-D363

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Park K, Kim MY, Vickers M, Park JS, Hyun Y, Okamoto T, Zilberman D, Fischer RL, Feng X, Choi Y, Scholten S (2016) DNA demethylation is initiated in the central cells of *Arabidopsis* and rice. Proc Natl Acad Sci U S A 113: 15138-15143

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Qiu Y, Kohler C (2020) Mobility connects: transposable elements wire new transcriptional networks by transferring transcription factor binding motifs. Biochem Soc Trans 48: 1005-1017

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43: e47

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Rodrigues JA, Hsieh PH, Ruan D, Nishimura T, Sharma MK, Sharma R, Ye X, Nguyen ND, Nijjar S, Ronald PC, Fischer RL, Zilberman D (2021) Divergence among rice cultivars reveals roles for transposition and epimutation in ongoing evolution of genomic imprinting. Proc Natl Acad Sci U S A 118

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Rodrigues JA, Ruan R, Nishimura T, Sharma MK, Sharma R, Ronald PC, Fischer RL, Zilberman D (2013) Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. Proc Natl Acad Sci U S A 110: 7934-7939

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Sabelli PA, Larkins BA (2009) The development of endosperm in grasses. Plant Physiol 149: 14-26

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Song X, Cao X (2017) Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. Curr Opin Plant Biol 36: 111-118

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Tonosaki K, Ono A, Kunisada M, Nishino M, Nagata H, Sakamoto S, Kijima ST, Furuumi H, Nonomura KI, Sato Y, Ohme-Takagi M, Endo M, Cornai L, Hatakeyama K, Kawakatsu T, Kinoshita T (2021) Mutation of the imprinted gene OsEMF2a induces autonomous endosperm development and delayed cellularization in rice. Plant Cell 33: 85-103

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Wells JN, Feschotte C (2020) A Field Guide to Eukaryotic Transposable Elements. Annu Rev Genet 54: 539-561

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Xing MQ, Zhang YJ, Zhou SR, Hu WY, Wu XT, Ye YJ, Wu XX, Xiao YP, Li X, Xue HW (2015) Global Analysis Reveals the Crucial Roles of DNA Methylation during Rice Seed Development. Plant Physiol 168: 1417-1432

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Yang G, Lee YH, Jiang Y, Shi X, Kertbundit S, Hall TC (2005) A two-edged role for the transposable element Kiddo in the rice ubiquitin2 promoter. Plant Cell 17: 1559-1568

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

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

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Yuan J, Chen S, Jiao W, Wang L, Wang L, Ye W, Lu J, Hong D, You S, Cheng Z, Yang DL, Chen ZJ (2017) Both maternally and paternally imprinted genes regulate seed development in rice. New Phytol 216: 373-387

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zemach A, Kim MY, Silva P, Rodrigues JA, Dotson B, Brooks MD, Zilberman D (2010) Local DNA hypomethylation activates genes in rice endosperm. Proc Natl Acad Sci U S A 107: 18729-18734

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zhang H, Lang Z, Zhu JK (2018) Dynamics and function of DNA methylation in plants. Nat Rev Mol Cell Biol 19: 489-506

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)

Zhang X, Jiang N, Feschotte C, Wessler SR (2004) PIF- and Pong-like transposable elements: distribution, evolution and relationship with Tourist-like miniature inverted-repeat transposable elements. Genetics 166: 971-986

Google Scholar: [Author Only](#) [Title Only](#) [Author and Title](#)