

RESEARCH ARTICLE

Transcriptional regulation of genes bearing intronic heterochromatin in the rice genome

Nino A. Espinas^{1,2}, Le Ngoc Tu¹, Leonardo Furci¹, Yasuka Shimajiri^{1,3}, Yoshiko Harukawa¹, Saori Miura¹, Shohei Takuno⁴, Hidetoshi Saze^{1*}

1 Plant Epigenetics Unit, Okinawa Institute of Science and Technology Graduate University, Onna-son, Okinawa, Japan, **2** Plant Immunity Research Group, RIKEN Center for Sustainable Resource Science (CSRS), Yokohama city, Kanagawa, Japan, **3** EditForce, Fukuoka, Japan, **4** Department of Evolutionary Studies of Biosystems, SOKENDAI (The Graduate University for Advanced Studies), Hayama, Kanagawa, Japan

* hidetoshi.saze@oist.jp

OPEN ACCESS

Citation: Espinas NA, Tu LN, Furci L, Shimajiri Y, Harukawa Y, Miura S, et al. (2020) Transcriptional regulation of genes bearing intronic heterochromatin in the rice genome. *PLoS Genet* 16(3): e1008637. <https://doi.org/10.1371/journal.pgen.1008637>

Editor: Ortrun Mittelsten Scheid, Gregor Mendel Institute of Molecular Plant Biology, AUSTRIA

Received: May 29, 2019

Accepted: January 28, 2020

Published: March 18, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pgen.1008637>

Copyright: © 2020 Espinas et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All the sequence data reported in this study have been deposited in the DDBJ Sequence Read Archive under accession ID

Abstract

Intronic regions of eukaryotic genomes accumulate many Transposable Elements (TEs). Intronic TEs often trigger the formation of transcriptionally repressive heterochromatin, even within transcription-permissive chromatin environments. Although TE-bearing introns are widely observed in eukaryotic genomes, their epigenetic states, impacts on gene regulation and function, and their contributions to genetic diversity and evolution, remain poorly understood. In this study, we investigated the genome-wide distribution of intronic TEs and their epigenetic states in the *Oryza sativa* genome, where TEs comprise 35% of the genome. We found that over 10% of rice genes contain intronic heterochromatin, most of which are associated with TEs and repetitive sequences. These heterochromatic introns are longer and highly enriched in promoter-proximal positions. On the other hand, introns also accumulate hypomethylated short TEs. Genes with heterochromatic introns are implicated in various biological functions. Transcription of genes bearing intronic heterochromatin is regulated by an epigenetic mechanism involving the conserved factor OsIBM2, mutation of which results in severe developmental and reproductive defects. Furthermore, we found that heterochromatic introns evolve rapidly compared to non-heterochromatic introns. Our study demonstrates that heterochromatin is a common epigenetic feature associated with actively transcribed genes in the rice genome.

Author summary

Intronic regions of eukaryotic genomes accumulate many Transposable Elements (TEs) and repeats. These intronic repeats are often targeted by epigenetic silencing mechanisms and form a repressive heterochromatin structure, even within transcriptionally active genes. However, the distribution of TEs in the intragenic regions, and their contributions to genetic diversity and evolution in plant genomes, remain poorly understood. In this study, we investigated the genome-wide distribution of intronic TEs and their epigenetic states in the *Oryza sativa* genome, where TEs comprise 35% of the genome. We found

DRA008322. All other data are within the manuscript and its Supporting Information files.

Funding: This work was supported by MEXT Grant-in-Aid for Scientific Research on Innovative Area (http://www.mext.go.jp/a_menu/shinkou/hojyo/1218181.htm) Grant Number 19H05272 to HS, and also supported by Okinawa Institute of Science and Technology Graduate University (<https://www.oist.jp>) to HS. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

that over 10% of rice genes contain introns associated with repressive heterochromatin. Genes with heterochromatic introns are implicated in various biological functions. The conserved protein OsIBM2 is required for proper transcription of a group of heterochromatin-containing genes. We also found that heterochromatic introns evolve rapidly compared to non-heterochromatic introns. Our study indicates that heterochromatin is a common feature in transcribed genes in the rice genome.

Introduction

Genomes of eukaryotes contain substantial numbers of transposable elements (TEs), which shape genomic structures and epigenomic landscapes [1, 2]. In plants, genomic TE contents are strongly correlated with genome size expansion [3]. Since TE insertions in genes disrupt coding sequences and regulatory elements, TEs are evolutionarily purged from genic regions and accumulated in the gene-poor pericentromeric regions of chromosomes, especially in species with small genomes, such as *Arabidopsis thaliana* [2, 4]. However, in plants with larger genomes, TEs are also distributed across the gene-rich chromosome arm regions, and often affect transcription of surrounding genes [2, 5–8].

Due to their harmful effects in the genome, TEs are often epigenetically modified and transcriptionally silenced by genome defense mechanisms [9, 10]. In plants, interdependent chromatin modifications including DNA methylation, histone modifications, and RNA interference (RNAi) play key roles in transcriptional repression of TEs. DNA cytosine methylation is found in both CG and non-CG (CHG, CHH; H = A, T, C) contexts in plant genomes, which is important in TE silencing. CG methylation is maintained through DNA replications by the Methyltransferase 1 (MET1) [11–14]. In addition, DNA methylation is directed by RNAi-based RNA-dependent DNA methylation, where small interfering RNAs (siRNAs) recruit *de novo* DNA methyltransferase to target sequences [15]. Furthermore, histone modifications, including histone H3 Lys9 methylation (H3K9me), are tightly linked to non-CG methylation, and are associated with repressive chromatin states [16]. Chromatin with these modifications results in the formation of a condensed repressive chromatin structure called heterochromatin [16–18], commonly associated with most TE sequences. The chromatin remodeler Decrease in DNA Methylation 1 (DDM1) is required for the maintenance of heterochromatin [19–21].

The formation of heterochromatin on TEs in genic regions causes transcriptional repression of surrounding genes in plant genomes [22, 23]. For example, heterochromatin associated with TEs and repetitive sequences in promoter regions often causes transcriptional repression of downstream genes [24–27]. Many TEs are also present in intronic regions especially in large plant genomes [28–31], likely due to less adverse effects on coding sequences compared to exonic insertions. Enigmatically, intronic TE sequences can also be targeted by repressive chromatin modifications, thus forming heterochromatic structure within transcription permissive chromatin environments [32]. In *Arabidopsis thaliana*, nuclear proteins, including INCREASE IN BONSAI METHYLATION 2 (IBM2)/ANTI-SILENCING1/SHOOT GROWTH1, are required for proper transcription of heterochromatin-containing genes [33–36]. IBM2 contains a Bromo-Adjacent Homology (BAH) domain and an RNA recognition motif, and in the *Arabidopsis ibm2* mutant, genes containing heterochromatic introns show a transcription defect due to premature termination of transcripts at the heterochromatic regions. Intronic heterochromatin tends to repress expression of associated genes in both animals and plants [37–42]. However, in some circumstances establishment and maintenance of

heterochromatin within intronic regions are critical for transcriptional control of the associated genes required for environmental responses and development [28, 43, 44]. For example, in *A. thaliana*, maintenance of H3K9 methylation and DNA methylation of intronic TEs is important for transcription of the RPP7 gene, which confers resistance against a plant pathogen [45]. In winter wheat, vernalization induces DNA hypermethylation of the intron of *VRN-A1*, which promotes expression of the gene [43]. On the other hand, in oil palm, loss of DNA methylation of an intronic TE arising during tissue culture alters the splicing pattern of the associated gene, resulting in a developmental abnormality of the fruit [46]. These observations suggest that heterochromatin formation in intragenic, especially intronic TEs, may have functional relevance to transcriptional regulation of associated genes, and would also profoundly influence on gene diversification and evolution. Indeed, plant introns often encode regulatory elements for recruitment of transcription factors that alter chromatin states, which lead to both transcriptional repression and activation of developmental genes [47–50]. However, epigenetic states of introns at a genome-wide scale, their impacts on gene regulation, functions of genes bearing intronic heterochromatin, and their contribution to genetic diversity in plant genomes are not well understood.

The *Oryza sativa* genome is an ideal model for investigating interactions between genes and TEs, since 35% of the genome consists of TEs that are widely distributed in genic regions [51–53]. In this study, we investigated the genome-wide distribution of intronic heterochromatin and its impact on transcriptional control of genes in the rice genome. We found that over 10% of rice genes contain introns associated with repressive heterochromatin, which are involved in various biological processes. Transcription of genes bearing intronic heterochromatin as well as other genes without heterochromatic introns are affected by a loss of function of the conserved factor OsIBM2, which is essential for development and reproduction of rice. Rapid evolution of heterochromatic introns suggests their potential impacts on the evolution of gene sequences.

Results

Accumulation of heterochromatic introns in the rice genome

To investigate the epigenetic states of intronic regions in the rice genome (*Oryza sativa* L. ssp. *japonica* CV. Nipponbare), we performed whole-genome bisulfite sequence (WGBS) analysis using the mature rice leaf tissue. We specifically focused on detecting repressive heterochromatic states in intronic regions. Non-CG DNA methylation, especially CHG methylation, is well correlated with the heterochromatic state of histone modifications such as H3K9 methylation in plant genomes [20, 31, 54, 55]. Therefore, CHG methylation level was used as a proxy to define a heterochromatic state of chromatin within intronic regions (see [methods](#) for more details). We identified 5,809 introns within 4,227 gene models that contain heterochromatic domains in the rice genome (Fig 1A, S1 and S2 Tables). This is about 11% of gene models in rice genome (IRGSP-1.0; 37,866 gene models), and is 10-fold more abundant than in *Arabidopsis thaliana* (S1A Fig). Heterochromatic introns accumulate CG, CHG, and CHH methylation, as well as H3K9 di-methylation (H3K9me₂; S1B Fig), similar to transposable elements (TEs; Fig 1B and 1C), indicating that they contain canonical heterochromatin. Loci containing heterochromatic introns are not biased toward repeat-rich pericentromeric regions, but are rather scattered throughout the rice chromosome arms (Fig 1D). RNA-seq analysis of the leaf tissue demonstrated that many of these loci are transcribed in the presence of intronic heterochromatin (Fig 1E), indicating that heterochromatic introns can co-exist within transcriptionally active genes in the rice genome.

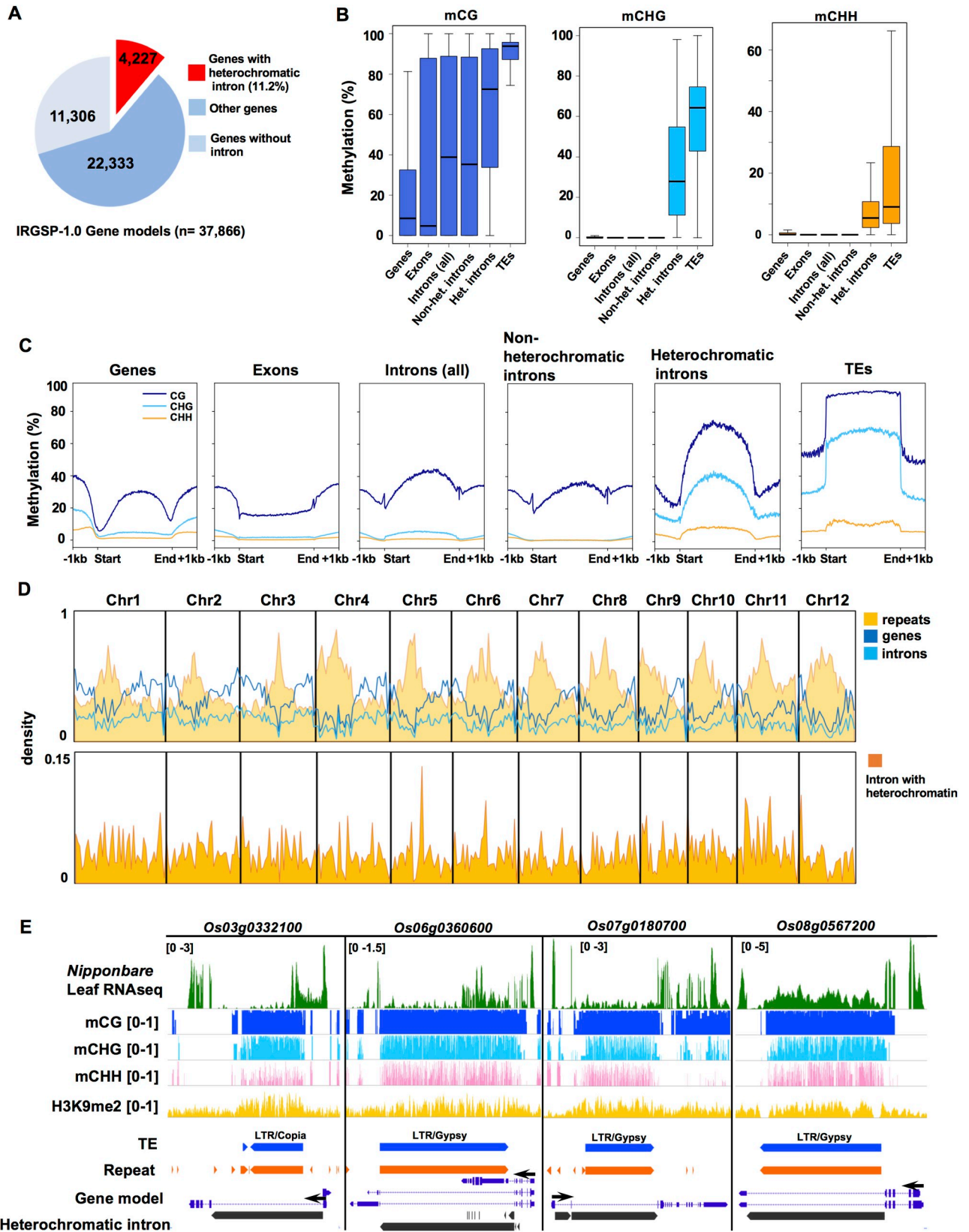


Fig 1. Intronic heterochromatin in the rice genome. (A) Rice IRGSP-1.0 gene models ($n = 37,866$) that contain heterochromatic domain in their intron ($n = 4,227$). (B) DNA methylation levels in CG (mCG), CHG (mCHG) and CHH (mCHH) contexts for indicated genome features. (C) Metaplots of DNA methylation in CG (blue), CHG (light blue) and CHH (orange) context for indicated genome features. (D) (Top) Density of repeats, genes, and intron sequences in 1MB bins in rice chromosomes. (Bottom) Density of introns with heterochromatic domains as above. (E) Representative rice genome loci containing heterochromatic domains within introns. Tracks: Top to bottom; RNAseq (Reads per Million are indicated at top left), mCG ratio (0 to 1), mCHG ratio (0 to 1), mCHH ratio (0 to 1), H3K9me2 (RPM; 0 to 1), TE annotation (blue), repeats (orange), gene model (purple), introns containing heterochromatic domain (black). Black arrows indicate the orientation of coding sequence.

<https://doi.org/10.1371/journal.pgen.1008637.g001>

Heterochromatin is enriched in promoter-proximal introns

In general, introns in the rice genome are longer than those in the *A. thaliana* genome (S2A Fig), which may be due to abundant repeat sequences in introns (13.9% of total intron sequence; S2B Fig). In particular, rice introns associated with CHG methylation tend to be longer (Fig 2A, S2C Fig) [34]. It has been reported that the first intron is generally longer than later introns in most of eukaryotic genomes, including that of rice [56] (Fig 2B). We found that heterochromatic introns are longer irrespective of their positions (Fig 2B). However, formation of heterochromatic introns is significantly biased toward the 5'-ends of rice genes ($p < 1.0e-6$ by a permutation test, Fig 2C), which is associated with accumulation of TEs in promoter-proximal introns (S2C Fig). This suggests that a preferential targeting of TEs toward the 5'-ends of rice genes might be a trigger for the formation of heterochromatin in promoter proximal introns.

Many intronic TEs are short and hypomethylated in CHG context

Next, we investigated how the presence of TE affects heterochromatin formation within intronic regions. A set of manually curated TE annotations ($n = 29,100$, S3 Table) was analyzed for their locations in the genome. We found that about 7% (2,122/29,100) of TEs are located within intragenic regions (Fig 3A, S3A Fig), and that most of them (82%; 1,751/2,122) are present in introns (Fig 3B). TEs annotated as Miniature inverted-repeat transposable elements (MITE) are particularly enriched in intragenic regions (S3A Fig) consistent with previous studies [57, 58], while no strong orientation bias against the associated genes was observed in any of the TE families (S3B Fig). As expected, most of heterochromatic introns (84%; 4,886/5,809) are associated with TEs and other repeat sequences (Fig 3C). Interestingly, however, about 50% (980/1,967) of TE-containing introns do not overlap with the heterochromatic introns (Fig 3C), suggesting that intronic TEs are not always associated with heterochromatin. On the other hand, 16% (923/5,809) of heterochromatic introns are not associated with TEs or with repeat annotations, which is likely due to a spreading of heterochromatic modifications from neighboring chromatin (S3C Fig) [59, 60]. DNA methylation of intronic TEs seems to be maintained in the same manner as intergenic TEs, since methylation in intronic TEs is affected by mutations of maintenance methylase OsMET1, and the chromatin remodeler OsDDM1 (S3D Fig) [12, 21]. However, we found that a fraction of intronic TEs is hypomethylated especially in CHG, while distribution of CG, and CHH methylation levels among TEs is comparable between intergenic and intronic TEs, irrespective of the TE families (Fig 3D, S4 Fig). CHG-hypomethylated TEs are generally shorter than CHG-hypermethylated TEs (S5 Fig), and they are more abundant in introns (Fig 3E). A similar trend was observed in an analysis using a comprehensive MITE dataset [61] (153,751 TE sequences), which showed that short, CHG-hypomethylated MITEs (S6A and S6B Fig) are enriched in introns. Shorter TEs are likely degenerated or truncated TE sequences, on which relaxed epigenetic silencing may have resulted in a reduction of CHG methylation.

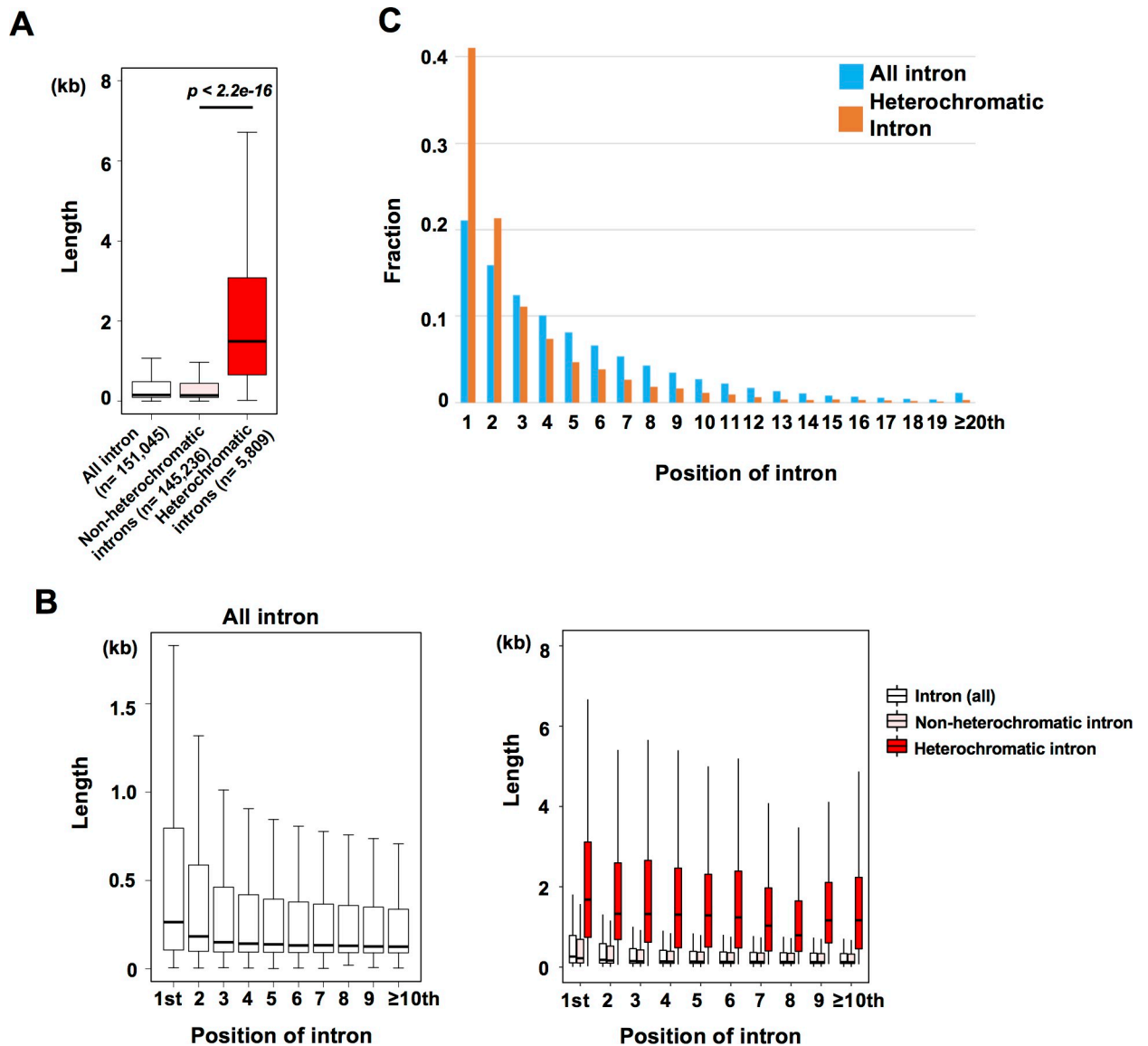


Fig 2. Heterochromatin is enriched in promoter-proximal introns. (A) Boxplots for length of normal and heterochromatic introns. Heterochromatic introns are significantly longer than introns without heterochromatic domains (p -value $< 2.2e-16$, Wilcoxon exact test). (B) (left) Intron position and length for all introns. (right) Intron position and length for all introns (white), non-heterochromatic introns (pink) and heterochromatic introns (red). (C) Enrichment of heterochromatin in promoter-proximal introns. Fraction of relative positions for all introns ($n = 151,045$), and heterochromatic introns ($n = 6,086$; the average position of heterochromatic introns was 3.02) are shown. Identical introns annotated in different positions in different splicing variants were independently counted.

<https://doi.org/10.1371/journal.pgen.1008637.g002>

Heterochromatic introns are associated with genes involved in various biological functions

Rice genes with heterochromatic introns encode various proteins with enzymatic activities, including oxidoreductases and hydrolases, as well as with nucleotide-binding activities (S7A Fig). Gene Ontology (GO) enrichment analysis indicates that genes with heterochromatic introns are implicated in diverse functions, such as lipid/carbohydrate metabolic processes, post-embryonic and reproductive developmental processes, and cell death pathway, which is a manifestation of plant defense responses against pathogens (Fig 4A). On the other hand, GO

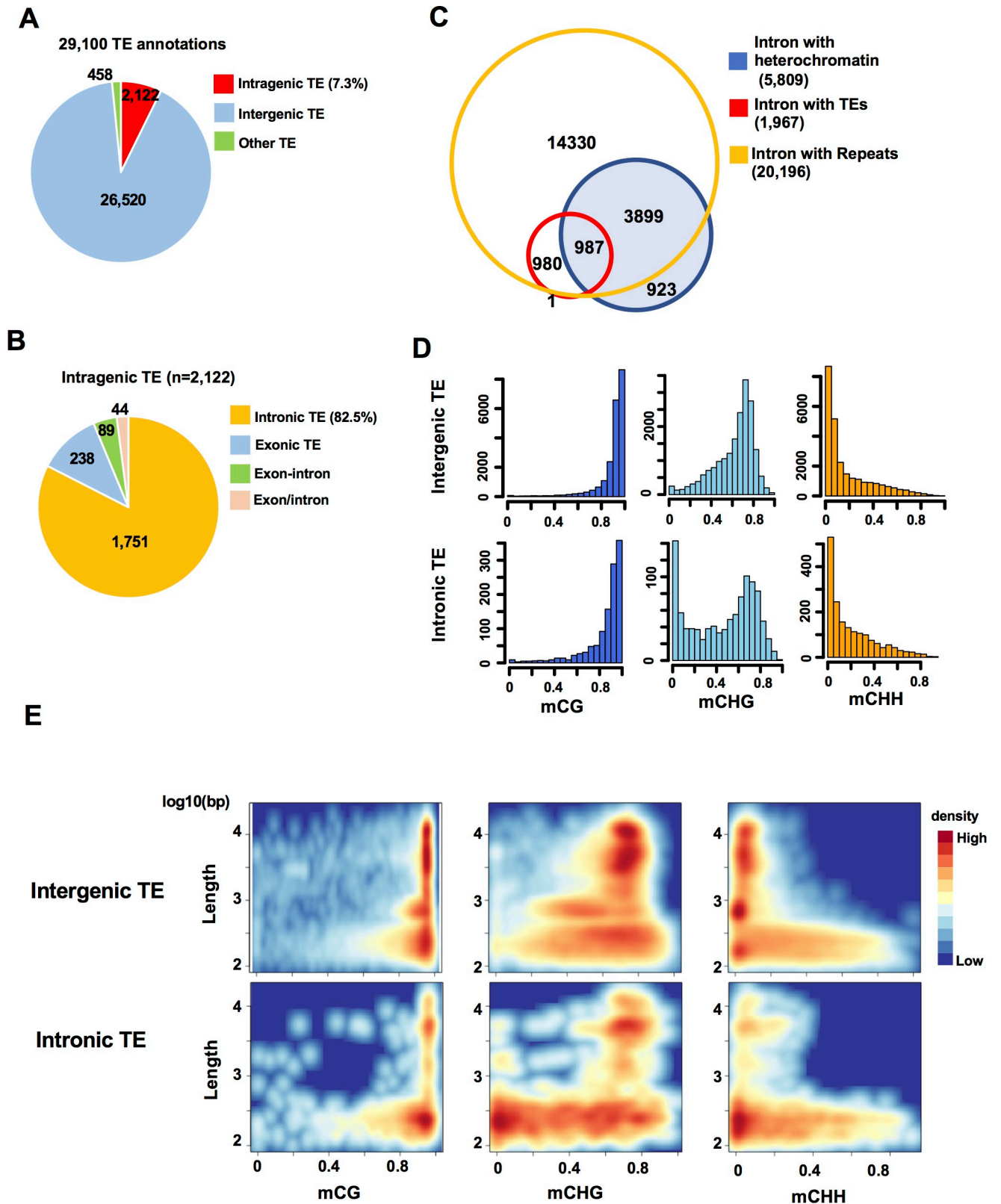


Fig 3. Intronic TEs are short and hypomethylated. (A) Classification of all TEs (n = 29,100) in the rice genome. “Other TEs” refers to TE annotations overlapping to both gene and intergenic regions. (B) Classification of intragenic TEs (n = 2,122) in the rice genome. “Exon-intron” refers to TE annotations

overlapping to both exon and intron. “Exon/intron” refers to TE annotations included in an exon of a gene/transcript model as well as in an intron of other gene/transcript models. (C) Venn diagram showing the number of overlapping introns containing heterochromatic domains (blue), TEs (red) and other repeats (yellow). (D) Histograms of the number of intergenic and intronic TEs and their methylation levels (0 to 1) in CG, CHG, and CHH contexts. TEs with methylation data at ≥ 5 Cs were included in the analysis. (E) Density plots showing length (\log_{10}) and methylation levels (0 to 1) of intergenic and intronic TEs in CG, CHG, and CHH contexts. TEs with methylation data at ≥ 5 Cs in each context were included in the analysis.

<https://doi.org/10.1371/journal.pgen.1008637.g003>

terms such as nitrogen biosynthetic/metabolic processes were depleted in the genes (S7B Fig). Our transcriptome analysis of mature rice leaf tissue showed that the expression levels of genes with heterochromatic introns were generally lower than those without heterochromatin (Fig 4B). We further examined expression patterns of rice genes in various developmental stages as well as in responses to environmental stimuli, using public microarray data in an expression atlas of rice genes and rice RNA-seq data [62, 63]. We calculated entropy values of gene expression patterns as a measure of specificity [64], which showed that genes with heterochromatic introns tend to have tissue-specific expression patterns, and are also responsive to plant hormones and environmental stresses (Fig 4C, S7C Fig). However, overall effect sizes of the values between gene with and without heterochromatic intron in the analyses were relatively small ($r < 0.1$), suggesting that expression profiles of genes with heterochromatic intron are not too different from genes without heterochromatin.

To understand the effects of heterochromatic introns on gene regulation in response to environmental signals, we searched for insertion/deletion polymorphisms in the intronic regions between Nipponbare (NB) and the indica-rice cultivar KASALATH (KAS) using whole-genome re-sequencing data [65]. In particular, we sought genes showing expression changes in response to JA, a plant hormone essential for development and also for both biotic and abiotic responses [66] (Fig 4C). Based on the genome re-sequencing data [67] and a public expression profile [63], we selected 12 JA-responsive loci (4 up-regulated, and 8 down-regulated loci after JA treatment; S8 and S9 Figs) that have large intronic deletions in the KASALATH genome corresponding to the regions showing heterochromatic state in the Nipponbare genome (S9 Fig). Consistent with the public expression profile, the JA-inducible genes *OsAOS2* [68, 69] as well as the 12 selected loci in NB showed expression changes in the root tissues upon JA treatment (S8 Fig). Several loci (4 out of 8 loci showing down-regulation in NB by JA treatment) in KAS showed reduced responses to JA treatment ($p > 0.05$; *t*-test), whereas other loci including up-regulated genes showed essentially similar responses between NB and KAS, with variable degrees (S8 Fig). Thus, impacts of heterochromatic intron on the gene response and expression remain to be elucidated.

A conserved epigenetic machinery regulates transcription of genes containing heterochromatin, and is essential for rice development

It has been shown in *A. thaliana* that transcription of genes with heterochromatic introns is regulated by a nuclear protein complex [33]. One of the proteins in the complex is Increased Bonsai Methylation 2 (IBM2), which contains a Bromo-Adjacent Homology (BAH) domain and an RNA recognition motif (S10 Fig)[34–36]. The rice homolog of IBM2 is encoded as a single-copy gene in the rice genome [34]. To examine whether it has a conserved function for transcription of genes with heterochromatic introns, we knocked down the transcript of the homologous gene (*Os01g0610300*; named as *OsIBM2*) using RNA interference (RNAi), targeting the 3' end of the gene (Fig 5A). Among several independent T1 transformants, lines #2 and #16 showed a marked reduction of the transcript and were further investigated (Fig 5B). In addition, the CRISPR-Cas9 system was employed to obtain *OsIBM2* knock-out lines, which generated independent deletion mutant lines targeting either the BAH domain-encoding

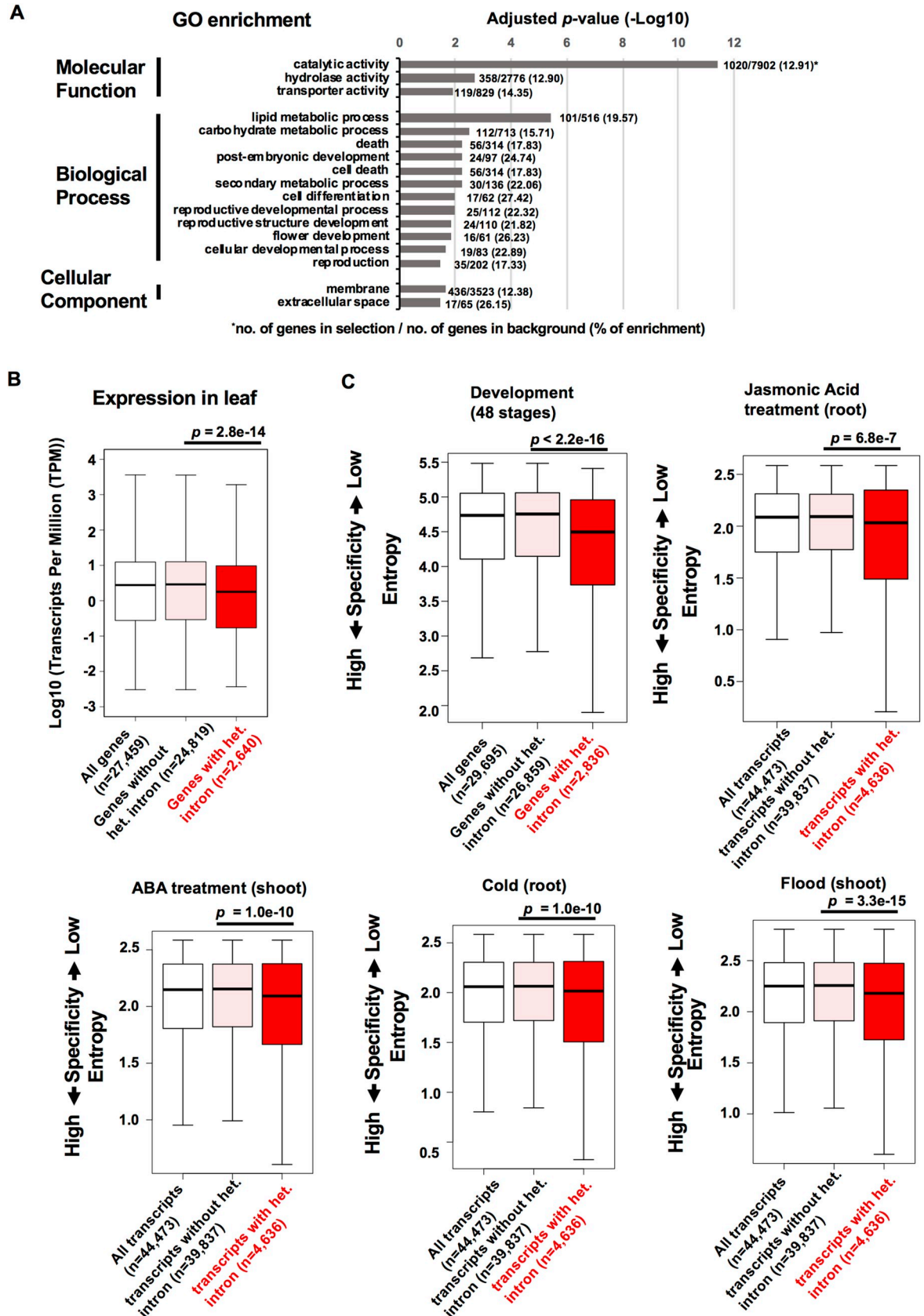


Fig 4. Genes with heterochromatic introns tend to show tissue-specific, and environment-responsive expression patterns. (A) Gene Ontology enrichment for genes containing heterochromatic introns (2,449 genes out of 4,227 genes were analyzed for enrichment analysis. p -values were obtained by Fisher test with Hochberg adjustments (FDR < 0.05). GO terms (odds ratio; 95% Confidence Interval): catalytic activity (1.43; 1.31, 1.56), hydrolase activity (1.29; 1.14, 1.46), transporter activity (1.44; 1.17, 1.76), lipid metabolic process (2.10; 1.67, 2.63), carbohydrate metabolic process (1.60; 1.29, 1.97), death (1.85; 1.36, 2.49), post-embryonic development (2.79; 1.68, 4.50), cell death (1.85; 1.36, 2.49), secondary metabolic process (2.40; 1.55, 3.66), cell differentiation (3.21; 1.72, 5.73), reproductive developmental process (2.44; 1.50, 3.86), reproductive structure development (2.37; 1.44, 3.78), flower development (3.02; 1.59, 5.46), cellular developmental process (2.52; 1.42, 4.27), reproduction (1.78; 1.19, 2.59), membrane (1.23; 1.10, 1.38), extracellular space (3.01; 1.62, 5.34). (B) Expression levels of genes with heterochromatic introns in the leaf tissue measured by RNA-seq (Transcript per million; TPM > 0). p -values by Wilcoxon test are indicated. Effect size $r = 0.046$. (C) Tissue specificity of normal genes (pink) and heterochromatin-containing genes (red) in the rice developmental process, and specificities for Jasmonic Acid (JA) and abscisic acid (ABA) treatments, and stress treatments (Cold, Flood), measured by entropy values. p -values from the Wilcoxon test are indicated. Effect size (r) in each analysis: Development; 0.095, JA; 0.024, ABA; 0.031, Cold; 0.031, Flood, 0.037.

<https://doi.org/10.1371/journal.pgen.1008637.g004>

region (*g1#5* and *g1#27*) or a 3' region downstream of the RRM encoding region (*g2#24*) (Fig 5C). Both RNAi and CRISPR-targeted mutants showed severe dwarfism and sterility (Fig 5D, S11A–S11C Fig). Particularly, mutants with deletions in the BAH domain (*g1#5* and *g1#27*) could not produce homozygous mutant seeds, suggesting an embryonic lethality of these alleles. Heterozygous mutants with a deletion in the 3' region (*osibm2_g2#24*) could produce homozygous seeds, but the homozygous plants showed a complete sterility (S11C Fig), indicating that *OsIBM2* is essential for development and reproduction. Previous studies in *Arabidopsis ibm2* have shown that transcription at downstream of heterochromatic introns is reduced due to a premature termination of transcript within heterochromatic introns [34]. Therefore, we analyzed changes in accumulation of transcripts upstream and downstream of introns with both heterochromatic and non-heterochromatic state (total of 126,068 introns) in the rice genome. Our transcriptome analysis of the leaf tissues from both RNAi and CRISPR-Cas9 mutant lines detected 454 differentially expressed genes (DEGs) commonly in RNAi_#2, #16 and *osibm2_g2#24* lines, which showed changes in transcripts downstream of introns compared with wild type (Fig 5E, S4 Table). Among DEGs, genes containing heterochromatic introns were significantly enriched (93 genes out of 454 DEGs (20.5%); $p = 2.9e-17$, Fisher's exact test). DEGs with heterochromatic introns showed a significant reduction of transcripts in the 3' downstream of the heterochromatic intron ($p = 1.0e-6$, Tukey-Kramer test; S12B Fig), which was due to premature polyadenylation in the intronic regions (Fig 5F and 5G, S13, S14A–S14C Figs), similar to the phenotypes of the *Arabidopsis ibm2* [34, 70]. On the other hand, DEGs with normal introns showed less changes in their 3' transcription (S12A and S12B Fig), suggesting that the mutation in *OsIBM2* results in transcription defects predominantly at heterochromatin-containing DEGs. We also searched for differentially expressed TEs in the *osibm2*. We detected only a few of them (23 TEs; 22 LTR, 1 DNA/En-Spm; 12 up-regulated, 11 down-regulated; S14D Fig), including 8 intronic TEs (3 TEs were associated with the DEGs containing heterochromatin; S14E Fig); some of these expression changes of TEs might be due to epigenetic changes during tissue culture transformation. The number of DEGs with and without heterochromatin that were detected by the RNA-seq analysis may have been underestimated, considering the partial loss of function of mutant alleles as well as the tissue-specific/environment-responsive expression profiles of heterochromatin-containing genes (Fig 4 and Fig 5). Indeed, additional RT-PCR analysis using RNAs from endosperm/embryo of *osibm2* showed that several heterochromatin-containing genes primarily expressed during reproductive development [71–76] were severely affected in *osibm2* (S11D Fig), even though they were not detected as DEGs in the RNA-seq of leaf tissues.

In *Arabidopsis*, the histone H3K9 demethylase gene, *IBM1*, contains heterochromatin in the 7th intron due to an insertion of organelle genome sequence, and *Arabidopsis ibm2* reduces expression of *IBM1*, which results in genome-wide accumulation of H3K9me2 and

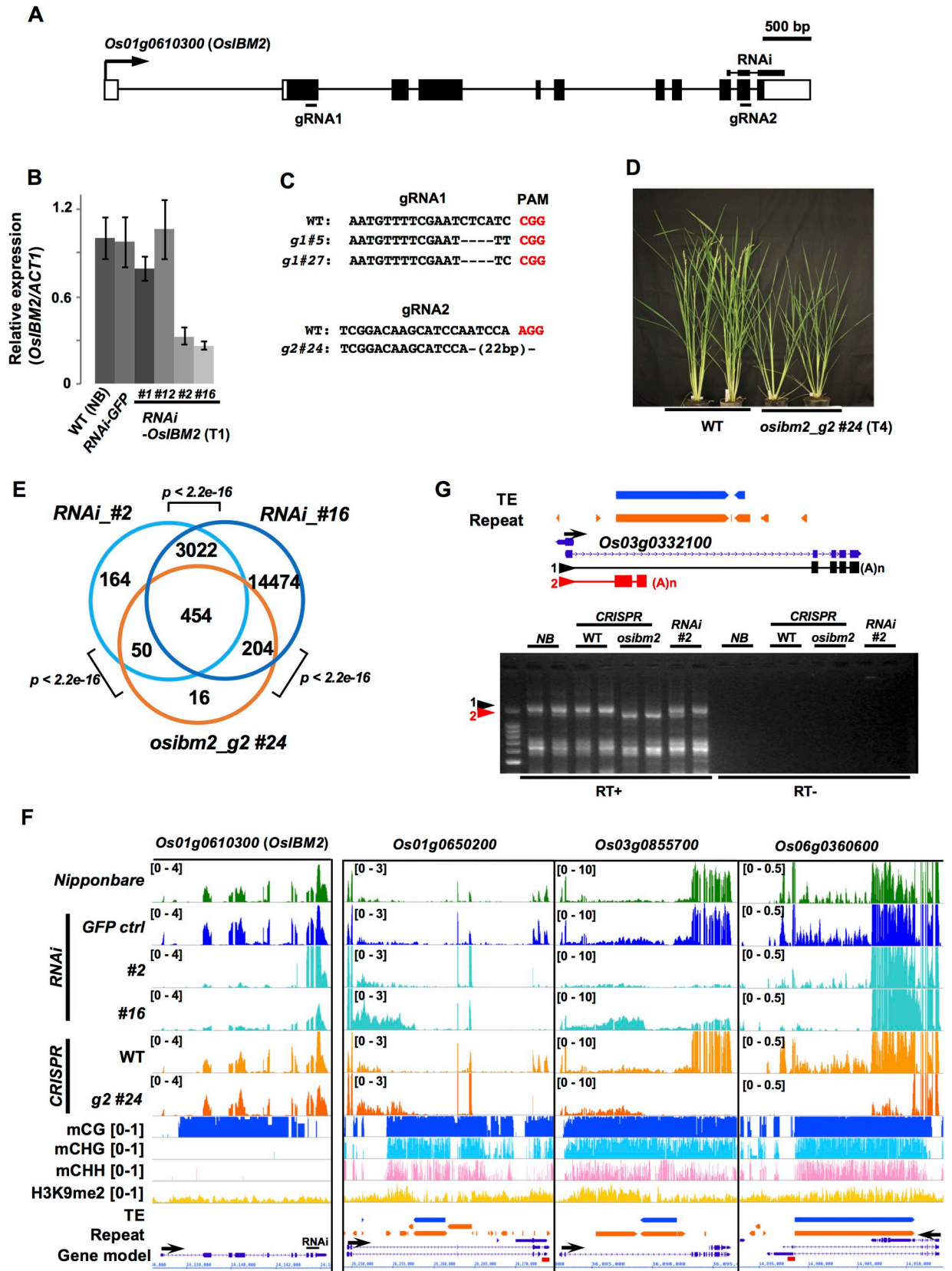


Fig 5. OsIBM2 is required for rice development and transcription of heterochromatin-bearing genes. (A) Gene structure of *Os01g0610300* (*OsIBM2*). Exons and untranslated regions are shown with black and white boxes, respectively. Regions designed for two gRNAs and hairpin RNA (RNAi) are also indicated. (B) qRT-PCR analysis of the expression of *OsIBM2* in 95-day-old leaf blade tissue of wild-type Nipponbare (NB), RNAi-*GFP* control line, and four RNAi-*IBM2* transgenic lines. Expression levels in each sample were normalized by *ACT1* expression levels, and the average of *OsIBM2/ACT1* in NB was set as 1. Bars represent means of three biological replicates \pm S.E.M ($n = 3$). (C) Cas9 gRNAs and targeted deletions obtained in independent *osibm2* mutants. PAM: Protospacer Adjacent Motif. (D) Three-month-old rice plants of *osibm2_g2#24* and their segregating wild type siblings (WT; T4). (E) Venn diagrams of overlapping genes showing altered expression in RNAi #2, #16, and *osibm2_g2#24*. *P*-values for significance of overlaps were tested with Fisher's exact test. (F) Representative rice genome loci showing altered expression patterns in mutants of *OsIBM2*. Tracks: Top to bottom; RNAseq (Reads per Million are indicated in top left), mCG ratio (0 to 1), mCHG ratio (0 to 1), mCHH ratio (0 to 1), H3K9me2 (RPM; 0 to 1), TE annotation (blue), repeats (orange), gene model (purple). The black arrow indicates the orientation of coding sequence. Red bars indicate primer positions for qPCR in S13B Fig. *OsIBM2* locus is shown as a validation of RNAi knock-down. (G) 3' Rapid Amplification of cDNA Ends (RACE) of *Os03g0332100* containing intronic heterochromatin. Upper panel: Structure of *Os03g0332100* locus and polyadenylated mRNA variants detected by 3' RACE. Exons and spliced introns confirmed by sequencing analysis are shown as black/red boxes and lines, respectively. Primer positions used for 3' RACE are indicated by arrows. Lower panel: Gel picture of DNA fragments amplified by 3' RACE. Two biological replicates for each genotype were examined. DNA fragments indicated by arrowheads were cloned and sequenced for at least 8 clones, and the representative sequences supported with more than 3 clones are shown in the upper panel. NB: Nipponbare; *osibm2*: *osibm2_g2#24*; WT: wild type segregants of *osibm2*; (A)n: polyadenylation.

<https://doi.org/10.1371/journal.pgen.1008637.g005>

non-CG methylation at genic regions [77, 78]. We therefore scrutinized, by WGBS analysis of *osibm2*, whether *OsIBM2* regulates non-CG methylation in the rice genome using the CRISPR mutant *osibm2_g#24*. We found that DNA methylation patterns in CG and non-CG contexts were nearly identical in genic as well as intergenic regions between *osibm2* and WT (S15 Fig). In the rice genome, two *IBM1* homologs, *OsJMJ718* (MSU ID: *Os09g22540*; RAP ID: *Os09g0393200*) and *OsJMJ719* (MSU ID: *Os02g01940*; RAP ID: *Os02g0109400*, *Os02g0109501*), have been identified (S16A and S16B Fig) [79]. We found that one of the *IBM1* homologs, *OsJMJ718* contains heterochromatin in the last intron (S16A Fig), although it was not identified as a commonly affected gene among *OsIBM2* mutants (significant transcript changes were detected in RNAi_#2 and #16; $q < 0.01$). The less significant effects of the *OsIBM2* mutation on *OsJMJ718* expression and genome-wide non-CG methylation may be due to a partial loss of function of *OsIBM2* in the mutants, or to functional redundancy of *OsJMJ718* and *OsJMJ719*. Alternatively, the *OsJMJ718* transcript may be more resistant to the effects of heterochromatin that is downstream of the *jmjC* domain-coding sequence, compared with the A.

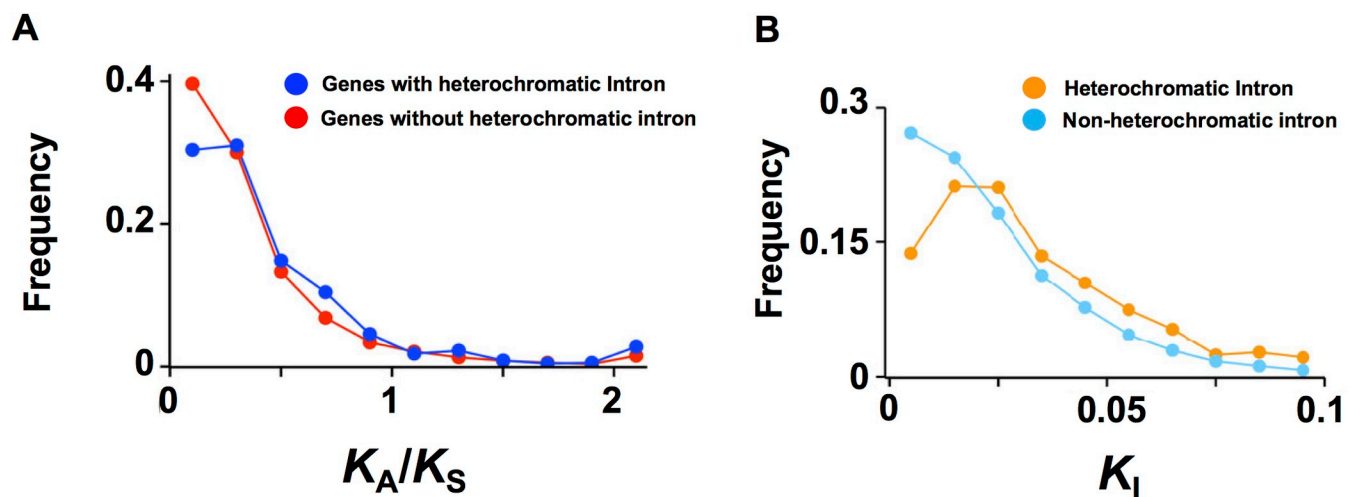


Fig 6. Patterns of nucleotide substitution rates between *Oryza sativa* and *Oryza meridionalis*. (A) Frequency distributions of K_A/K_S values. Blue and red plots represent genes with and without heterochromatic introns, respectively. (B) Frequency distributions of K_I values. Orange and light blue plots represent heterochromatic introns and heterochromatin-free introns, respectively.

<https://doi.org/10.1371/journal.pgen.1008637.g006>

thaliana IBM1, which has the heterochromatic intron in the middle of jmjC domain-coding sequence (S16C Fig).

Rapid evolution of heterochromatic intron

To understand how heterochromatin formation affects gene evolution in the rice genome, we further investigated the pattern of nucleotide substitutions in rice genes with heterochromatic introns. We first tested whether the degrees of selective constraints are similar between genes with and without heterochromatic introns in *O. sativa*. To this end, we compared the genome sequence of *O. sativa* with that of a close wild relative, *O. meridionalis* [80]. We predicted the orthologs in *O. meridionalis* and calculated the rate of nucleotide substitutions (Materials and Methods). Although our previous study of the *A. thaliana* genome did not find a significant difference [28], we found a relaxation of selective constraints in genes with heterochromatic introns in the rice genome (Fig 6A), where the ratio of nonsynonymous substitution rates to synonymous substitution rates (K_A/K_S) was 0.473 ($n = 928$), compared to 0.384 ($n = 10,456$) in genes without heterochromatic introns ($P < 10^{-5}$ by a permutation test). This indicates that heterochromatic introns would be deleterious for genes under high levels of selective constraint.

We further investigated the pattern of nucleotide substitutions in introns (K_I). Even though we excluded repeat sequences for the inter-species comparison, K_I values of heterochromatic introns showed higher base substitution rates (0.0325; $n = 627$) than non-heterochromatic introns (0.0242; $n = 35,354$; $P < 10^{-5}$ by a permutation test) (Fig 6B). This indicates that heterochromatic introns have evolved more rapidly than heterochromatin-free introns, suggesting an acceleration of intronic sequence divergence associated with heterochromatin formation.

Discussion

In this study, we revealed the genome-wide distribution of heterochromatic introns in the rice genome, which contains heterochromatic introns in approximately 11% of the genes. The underlying molecular mechanisms that allow the presence of repressive heterochromatin within actively transcribed regions are still unclear [37]. However, our study demonstrated that the conserved epigenetic factor OsIBM2 is critical for production of proper mRNA through heterochromatic introns in dozens of loci in the rice genome (Fig 5, S11D Fig). In addition, many genes without heterochromatic introns are also affected by OsIBM2 mutation in the leaf tissue (Fig 5), suggesting the profound impact of the loss of function of OsIBM2. Rice mutants of major epigenetic regulators, including OsMET1, OsCMT3, OsDRM and OsDDM1, have been shown to exhibit severe developmental defects such as embryonic/seedling lethality and sterility [14, 20, 81–83], while phenotypes of mutants of these genes in *A. thaliana* are relatively mild and plants are essentially viable [11, 84–87]. The difference likely stems from the genome structure of rice, where abundant TEs are distributed along gene-rich chromosome arms [51]. The close association of TEs with genes would make genes more susceptible to epigenetic changes in nearby TEs. In *A. thaliana*, *ibm2* plants are still fertile [34], while rice *osibm2* results in severe developmental defects and sterility (Fig 5, S11 Fig). This suggests that in the rice genome, in addition to the maintenance of heterochromatic states by DNA methyltransferases, transcriptional regulation of genes by OsIBM2 affects rice development and reproduction. For plant genomes harboring abundant intragenic heterochromatin, gene regulation mechanisms involving IBM2 would be more vital [29–31].

Insertion of TEs in intronic regions often results in repression of associated genes due to accumulation of repressive epigenetic marks. In rice, insertions of MITE in an intron cause

repression of the Elongated Uppermost Internode (EUI) gene, which is due to siRNA production from the intronic TEs [88]. In *Arabidopsis* and *Capsella* natural strains, insertions of TEs in the intron of flowering repressor gene *Flowering Locus C (FLC)* downregulate its expression, and induce early-flowering phenotypes [41, 59, 89]. Consistent with these reports, we found that rice genes with heterochromatic introns tend to show lower expression in the leaf tissue (Fig 4B). Alternatively, genes expressed at lower levels may tolerate those insertions [90]. On the other hand, our analysis of JA-responsive loci with insertion/deletion polymorphisms in heterochromatic introns suggested that responsiveness of genes to the hormone are largely unaffected by the presence/absence of heterochromatin in introns (S8 and S9 Figs). However, further comprehensive analyses are required to fully understand the impacts of intronic heterochromatin on gene regulation during environmental responses.

Longer first introns are a universal feature of eukaryotic gene structure [56]. The first intron sequence is more conserved than the later introns in animal genomes [91, 92]. In plants, enhancement of gene expression by intronic sequences, known as intron-mediated enhancement (IME), is associated with specific sequence motifs enriched in the first intron [93]. In rice, the first introns are required for the higher expression of tubulin genes [94, 95]. Intriguingly, intronic heterochromatin is significantly enriched in first and second introns which are associated with the accumulation of TEs in these introns (Fig 2, S2C Fig). Many TEs are known to target the 5' end of genes [7, 96, 97], while insertions into the exons in the 5' ends of genes would be selected against, which may result in the accumulation of TEs in promoter-proximal introns. Insertion of TEs and formation of repressive chromatin may physically disrupt or override transcription enhancer functions of the promoter-proximal introns, which may contribute to lower expression of the associated genes (Fig 4B). Additionally, the inserted TEs may provide novel regulatory sequences such as transcription factor binding sites [6, 22], allowing genes to acquire tissue-specific, or environment-responsive expression properties (Fig 4). The degree of selective constraints and tissue specificity are negatively correlated in *Arabidopsis* species [98]. Consistent with this, we observed that genes with heterochromatic introns tend to be expressed in a tissue-specific manner, and to show a lower degree of selective constraints than the other genes (Fig 6). We also observed that heterochromatic intron sequences show higher evolution rates (Fig 6), likely due to the higher mutation rates of methylated cytosine residues [99]. Thus, the formation of heterochromatin in intronic regions may contribute to the divergence of gene sequences.

The association of repetitive elements with genes is most prominent in disease-resistance gene (R gene) loci in plant genomes, which would accelerate gene diversification by enhancing recombination, and by shuffling and duplication of the sequences [6]. Indeed, R-genes are significantly overrepresented in genes with heterochromatic introns (119 out of 689 R-genes; 2.8% of 4,227 genes with heterochromatic introns; $p = 1.66e-6$, Fisher's exact test). Also, our GO analysis showed that heterochromatic introns are enriched in genes involved in the cell death pathway, which is provoked during plant immune responses mediated by R-genes [100, 101]. Acquiring repressive chromatin by TE insertions within intronic regions may also contribute to reduced expression of R genes, which may be advantageous for the prevention of autoimmune responses in the absence of pathogens [102].

A recent study showed that wild rice genomes tend to accumulate TEs in genic regions, while cultivated rice genomes show depletion of TEs from genic regions including introns [103]. This has likely occurred independently in the genomes of several cultivars [103]. This convergent loss of genic TE sequences in cultivar genomes may be a result of selective pressure against long heterochromatic TEs in the genic regions during domestication and selection (Fig 3). Alternatively, under uniform growing conditions in a nutrition-rich environment, inbreeding cultivar genomes may have gradually lost environment-responsive regulatory elements

associated with genic TEs. In contrast, longer introns with TE insertions in wild rice genomes may be adaptive for dynamic transcription changes in the fluctuating natural environment. Indeed, recent studies in budding yeast demonstrated that the presence of introns promotes survival under starvation conditions, while the introns are dispensable in a nutrient-rich environment [104, 105]. Intron sequences in plant genomes may have more profound impacts on genome evolution and plant adaptation than previously thought.

Methods

Rice genome annotations

Annotations of *Oryza sativa* genome, version IRGSP v1.0, locus/transcript/repeat annotations (IRGSP-1.0_representative_2015-03-31_2) were retrieved from RAP-DB (<http://rapdb.dna.affrc.go.jp/>) [106]. We identified TEs in the *Japonica* rice genome using RepeatMasker (ver. 4.0.5; <http://www.repeatmasker.org>). Repbase library (ver. 20140131) [107] was downloaded and used as a repeat library. We ran RepeatMasker with the default parameters and screened putative TE segments. We first excluded non-TE repeats such as simple repeats, rRNAs and satellite DNAs. We then further filtered out the following results; 1) the hit regions covering <70% of the total length of the repeats in the library, 2) the length of the hit regions is < 100 bp, 3) nucleotide divergence between the hit region and the repeat in the library is >20%. The list of TEs is in S3 Table. MITE annotation was retrieved from the P-MITE database [61], and used for a BLASTN [108] search of the IRGSP genome with a cutoff e-value of 1e-40. MITE sequences with identical lengths to query sequences having no mismatch and no gap (153,751 sequences) were used for further analysis. Chip-seq data for H3K9me2 was obtained from [109]. BS-seq data for *osmet1* and *osddm1* were obtained from [12, 21], respectively. Rice seed core collections (World Rice Core Collection; WRC) were obtained from Genebank Project, National Agriculture and Food Research Organization (NARO; https://www.gene.affrc.go.jp/databases-core_collections_wr.php).

Rice transgenic lines

All rice plants used in this study were grown in growth chambers under short-day condition (10 hours light/ 14 hours dark cycles) at 30°C during daytime and 25°C during the night. For RNAi knock-down of the *OsIBM2* mRNA, about 500 bp of the cDNA sequence of *OsIBM2* was cloned into pANDA vector [110]. A partial GFP sequence was used as a control RNAi vector. Wild-type Nipponbare calli were transformed with the RNAi vector at InPlanta Innovations (Yokohama, Japan) or at our laboratory, and more than 15 independent T1 transformants for each vector were obtained. For CRISPR-Cas9 knock-out of *OsIBM2*, two guide RNAs (S5 Table) were designed and cloned into pHUE411 (Addgene #62203) by GoldenGate Mix (NEB), and transformed into rice calli with a standard agrobacterium transformation method. Gene targeting events were detected by digestion with HpaII (gRNA1) or HaeIII (gRNA2), and were confirmed by Sanger sequencing. For *osibm2_g2#24*, the absence of the pHUE411 vector and fixation of the mutation (Fig 5C) were confirmed at T3. Segregating wild type (WT) and homozygous T4 plants were used for further analyses.

All oligonucleotides used in this study are listed in S5 Table.

Bisulfite sequencing and data analysis

For Whole Genome Bisulfite-Sequencing (WGBS) analyses, we used genomic DNA of Nipponbare, *osibm2_g2#24* (T4), and wild-type segregants of *osibm2_g2#24* (T4) isolated from three-month-old mature leaf tissues with Nucleon PhytoPure (GE). An Illumina Sequencing

libraries (125 bp paired-end for Nipponbare, 150 bp paired-end for *osibm2_g2#24* and Wild-type) were constructed using the PBAT method [111] and sequenced at OIST Sequencing Center (SQC). Raw reads were trimmed by Trimmomatic [112] with parameters; HEAD-CROP:10 SLIDINGWINDOW:4:20 MINLEN:50. Remaining paired reads were mapped to rice genome IRGSP v1.0 with Bismark (v0.19.0) [113] with parameters; -N 1—pbat -ambiguous -R 10 -un—score_min L,0,-0.6. Unmapped reads together with dropped single-end reads from trimming were further mapped to the rice genome as single-end reads with parameters; -N 1—pbat -ambiguous -R 10—score_min L,0,-0.6, for R1, and -N 1 -ambiguous -R 10 -un—score_min L,0,-0.6, for R2. Methylation reports from paired and single reads were merged with bedtools [114]. Only uniquely mapped reads were used for further analysis, and C bases covered by fewer than 3 reads, and also Cs more than 100 (Cs with unnaturally high coverage; top ~0.01% of covered Cs) were excluded. Methylcytosines were identified by binomial test [115], with the bisulfite conversion rate estimated by mapping sequencing reads to the rice chloroplast genome. Methylation levels were calculated using the ratio of $\#C/(\#C + \#T)$ as described in [116]. Methylcytosine domain containing consecutive ≥ 5 mCHG with ≥ 0.5 methylation on average was considered as heterochromatic domain. Boxplots, sequence density, and metaplots for DNA methylation were generated with deeptools [117], Microsoft Excel, and R. A summary of WGBS analyses is shown in [S6 Table](#).

GO analysis

GO analysis for enrichment was performed using the AgriGO website [118] and significant terms were extracted by Fisher's exact test with Hochberg adjustment (FDR<0.05). GO term depletion analysis was performed with TopGO (<https://rdr.io/bioc/topGO/>) using Fisher's exact test. Protein classes were determined using the Panther database [119].

Expression data analysis

Micro-array data and RNA-seq data were retrieved from RiceXpro database [62] and TENOR [63]. For gene expression of developmental stages, gene expression profiles of 48 rice developmental stages/tissues were used for calculation of entropy value of each gene. For gene expression profiles of stress/hormone treatment conditions, gene expression data at following time points were used for calculation of entropy value of each gene: Jasmonic Acid, ABA, Cold, Drought treatments; 0, 1, 3, 6, 12, 24 hours, Flood treatment; 0, 1, 3, 6, 12, 24, 72 hours, Osmotic stress; 0, 1, 3, 6, 12 hours, High/Low phosphate treatments; 0, 1, 5, 10 days. Entropy (modified H) was calculated with ROKU function in TCC package in R [120].

Genome sequencing data analysis for indel identification

Genome resequencing data of KASALATH genomes was retrieved from [65] and mapped to IRGSP-1.0 using bowtie2-2.2.2 [121]. Candidate loci for intronic deletion in the KASALATH genome were searched based on the INDEL data retrieved from [122]. The presence of deletion was confirmed by PCR.

Jasmonic Acid treatment and gene expression analysis

The rice strains, Nipponbare and KASALATH (WRC 2), were germinated on plates. After 6 days, seedlings were transferred to 15 mL plastic tubes and grown hydroponically in 1/10 Murashige-Skoog (MS) media for 4 days in a growth chamber as described above. Plants were transferred to 1/10 MS media containing the final 100 μ M Jasmonic Acid (JA, SIGMA) and 0.02% DMSO, or 0.02% DMSO as a mock treatment. After 6 hours of treatments, total RNA

was extracted from the roots using Maxwell 16 LEV Plat RNA kit (Promega), and Quantitative RT-PCR (qRT-PCR) was performed for analysis of gene expression.

RNA-seq analysis

For RNA-seq analysis, total RNA from the leaf tissues was isolated with Maxwell 16 LEV Plat RNA kit (Promega). Two biological replicates for Nipponbare (NB), GFP-RNAi control lines (T2), RNAi #2 lines (T2), *osibm2_g2#24* lines (T4), and wild-type segregants of *osibm2_g2#24* (WT; T4) were prepared. An additional NB line was used for comparison with the single RNAi #16 (T2) line. Illumina RNA-Seq libraries (150bp paired-end) were prepared and sequenced at the OIST Sequencing Center. Raw reads were trimmed with Trimmomatic with the following parameters; HEADCROP:10 LEADING:15 TRAILING:15 SLIDINGWINDOW:10:15 MIN-LEN:25. Remaining paired reads were mapped to rice genome IRGSP v1.0 with Hisat2 [123] with parameters;—min-intronlen 20—max-intronlen 20000. Exon and splicing junction information was specified by the annotation retrieved from RAP-DB to prepare a genome index for Hisat2. A summary of RNA-seq analysis is shown in S7 Table. Reads mapped to rDNA and tRNA were removed with bedtools. For visualization of RNA-seq read tracks, read duplication was removed with samtools [124], and Reads per million (RPM) for 1 bp bin was calculated with deeptools. The read tracks were visualized in Integrated Genome Browser [125]. For estimation of expression level, reads mapped on transcript annotations were counted with the featureCounts function in Rsubread [126] with parameters; allowMultiOverlap = TRUE, min-Overlap = 1, fracOverlap = 0, countMultiMappingReads = FALSE, and used for Transcript Per Million (TPM) calculations for each gene model (Fig 4B). Expression changes in RNAi (RNAi_#2 and _#16) and CRISPR knock-out lines (*osibm2_g2#24*) were analyzed based on the methods in [34]. Transcripts mapped to pre- and post- introns ($n = 126,068$) in each gene model were counted by featureCounts. Ratio of read counts (mapped reads in pre-intron: mapped reads in post-intron) of two biological replicates of each genotype (RNAi_#2 lines vs RNAi_GFP lines, *osibm2-g2#24* lines (T4) vs wild-type segregants of *osibm2-g2#24* (WT; T4)) were tested to detect changes in the expression pattern, by employing logistic regression analysis with p -value correction by Benjamini-Hochberg (BH) method for multiple testing. Changes in gene expression between RNAi_#16 (one replicate) and control NB were detected by binominal test with p -value correction as above. Data sets with $q \leq 0.01$ were considered as significantly changed in downstream transcription (both up- and down-regulated loci in 3' region). A relative 5'/3' ratio of transcripts mapped to up- and down-stream of introns was calculated as described previously [34]. Differential expression analysis of TEs was performed by DESeq2 [127] using mapped read data by Hisat2 (*osibm2-g2#24* lines (T4) vs wild-type segregants of *osibm2-g2#24* (WT; T4)).

Quantitative RT-PCR (qRT-PCR) and 3' RACE were performed as described in [34].

Nucleotide substitution analysis

To reveal patterns of nucleotide substitutions in genes with heterochromatic introns, we compared nucleotide sequences of *O. sativa* and *O. meridionalis* [80]. Putative orthologs were identified using GenomeThreader [128] with mRNAs of *O. sativa* to find orthologs in *O. meridionalis*, with the following parameters; -minmatchlen 18 -seedlength 16 -exdrop 2. When multiple orthologs were detected for an mRNA, it was discarded. If no ortholog was detected, we incremented the parameter -exdrop by one. This process was repeated until a single ortholog was detected or until the parameter -exdrop was less than or equal to 5. We further screened orthologs in which the exon-intron structures were conserved between the orthologs in 80% of their nucleotide sequences after alignments with CLUSTALW2 [129]. Nonsynonymous and synonymous

nucleotide substitution rates (K_A and K_S , respectively) were calculated using the Nei and Gojori method [130]. We discarded genes with $K_S > 0.1$. We also calculated nucleotide substitution rates in introns as p -distance.

Supporting information

S1 Fig. Heterochromatic introns in *Arabidopsis thaliana* and rice genomes. (A) *Arabidopsis thaliana* genes (TAIR10) containing intron with heterochromatic domains. (B) Heatmap showing accumulation of H3K9 di-methylation on genome features in the rice genome. Data from [109] were used for the analysis.

(PDF)

S2 Fig. Length of introns in *Arabidopsis thaliana* and rice genomes. (A) A comparison of intron length between *Arabidopsis thaliana* ($n = 127,836$; average 169.0 bp) and *Oryza sativa* ($n = 126,068$; average 446.9 bp). (B) Fraction of repetitive elements in intronic regions of the rice genome. (C) Enrichment of heterochromatin and TEs in promoter-proximal introns. Fractions of all intron ($n = 151,045$), and heterochromatic introns ($n = 6,086$), and TE-containing introns ($n = 1,982$) are shown in the relative positions. Identical intronic regions annotated in different positions in different splicing variants were independently counted.

(PDF)

S3 Fig. TE families in rice introns. (A) Fraction of TE families in the intronic regions of the *Oryza sativa* genome. (B) Orientation of intronic TE insertion against gene annotations in each TE family. No significant orientation bias was observed in the TE families ($p > 0.01$; two-sided binominal test). (C) Metaplots of DNA methylation in CG, CHG and CHH contexts for heterochromatic introns with TEs and repeats ($n = 4,886$), heterochromatic introns without repeat ($n = 923$), and non-heterochromatic introns ($n = 145,235$). (D) Heatmap of methylation profiles of intronic TEs in wild-type *O. sativa* and mutants of *OsMET1* (*met1*) and of *OsDDM1* (*ddm1*) at CG, CHG, and CHH-contexts.

(PDF)

S4 Fig. DNA methylation of rice intergenic and intronic TEs. Histograms of the number of representative intergenic and intronic TE families (>20 copies in each category) and their methylation levels (0 to 1) in CG, CHG, and CHH contexts. TEs with methylation data at ≥ 5 Cs were analyzed.

(PDF)

S5 Fig. Length and DNA methylation of intronic TEs. Boxplots showing length of representative intergenic and intronic TE families (>10 copies in each category) and their methylation levels in CG (high; $mCG \geq 0.9$, low; $mCG < 0.9$), CHG (high; $mCHG \geq 0.2$, low; $mCHG < 0.2$), and CHH (high; $mCHH \geq 0.1$, low; $mCHH < 0.1$). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Wilcoxon exact test. N.S.: no significance, $p \geq 0.05$. TEs with methylation data at ≥ 5 Cs were analyzed.

(PDF)

S6 Fig. DNA methylation of MITEs in rice introns. (A) Histograms of the number of representative intergenic and intronic MITEs (data retrieved from the P-MITE database [61]) and their methylation levels (0 to 1) in CG, CHG, and CHH contexts. TEs with methylation data at ≥ 5 Cs were used in the analysis. (B) Density plots showing length (\log_{10}) and methylation levels (0 to 1) of intergenic and intronic MITEs in CG, CHG, and CHH contexts.

(PDF)

S7 Fig. Protein classes and expression changes of genes containing heterochromatic introns. (A) Protein classes defined by the Panther database [119]. 1,407 of 4,227 genes containing heterochromatic introns matching the database are indicated. (B) Gene Ontology depletion for genes containing heterochromatic introns. *P*-values were obtained by Fisher test, and terms with FDR < 0.05 are indicated. (C) Expression changes of all genes and genes with or without heterochromatic introns by various stress treatments. Specificity of the responses to given treatments were measured as entropy values. *P*-values from Wilcoxon exact test are indicated. Effect size (*r*) in each analysis: Low phosphate; 0.024, High phosphate; 0.020, Drought; 0.007, Osmotic stress; 0.009.

(PDF)

S8 Fig. JA response of genes in Nipponbare (NB) and KASALATH (KAS) with structural variations in heterochromatic intron. (A) Heatmap showing expression levels of the indicated genes after Jasmonic Acid (JA) treatment in the Nipponbare root. Expression data were obtained from TENOR [63]. (B) Quantitative RT-PCR (qRT-PCR) analysis of genes before (pre-treatment), and after JA (JA treatment). *OsAOS2* was included as a control for JA-dependent induction of expression. Relative expression levels in each sample were normalized by *UBQ1* expression levels, and the average of expression values in pre-treatment NB samples was set as 1, and plotted as dots (*n* = 6) with blue (NB) and yellow (KAS). The large dots and bars represent means of 6 biological replicates ± standard deviation (S. D.). *P*-values were obtained by *t*-test.

(PDF)

S9 Fig. Structural variations of heterochromatic introns in Nipponbare and KASALATH strains. Insertion/deletion polymorphisms in Nipponbare and KASALATH. Tracks: Top to bottom: mCG ratio (0 to 1), mCHG ratio (0 to 1), mCHH ratio (0 to 1), genome-resequencing data coverage (0 to 30) [65], repeats (orange), TE annotation (blue), gene model (purple). Structural variations detected by PCR are indicated under the tracks as gel pictures. Red arrows indicate the primer positions used for PCR amplifications shown in the gel panel. The region used for qRT-PCR is indicated as red bar.

(PDF)

S10 Fig. Amino acid alignment of homologs of OsIBM2. Amino acid alignment of homologs of OsIBM2 in plants based on [34]. Bromo-Adjacent Homology (BAH) domain and RNA-Recognition Motif (RRM) are framed with a blue line. Arrows indicate regions designed for guide RNAs used for CRISPR-Cas9 mediated deletion. At; *Arabidopsis thaliana*: Zm; *Zea mays*: Os; *Oryza sativa*: Sb; *Sorghum bicolor*: Pt; *Populus trichocarpa*: Rc; *Ricinus communis*.

(PDF)

S11 Fig. Developmental phenotypes of *osibm2* mutants. (A) Whole plant picture of three-month-old Nipponbare (left), RNAi_#2 line (middle) and RNAi_GFP control line (right). (B) Close-up pictures of seeds set in Nipponbare and RNAi lines (T1). (C) A close-up picture of seeds set in *osibm2_g2#24* and their segregating wild-type siblings (WT; T4). White bar: 1 cm. (D) RT-PCR analysis of gene expression in endosperm and embryo of Nipponbare and *osibm2*. RNAs from ~10 DAF (Days After Fertilization) developing endosperm and embryo of *osibm2_g2 #24* (T2) were used for the analysis.

(PDF)

S12 Fig. Expression changes in genes containing heterochromatic introns in *osibm2*. (A) (Top) DNA methylation levels of differentially expressed genes (DEGs) with heterochromatic introns (*n* = 93), DEGs without heterochromatic intron (*n* = 361), and non DEGs (*n* = 20293)

in Nipponbare background. (middle) DNA methylation difference in *osibm2* (*osibm2_g2 #24*) and wild type at loci as above. (Bottom) H3K9 methylation levels at loci as above. (B) 5'/3' ratio of transcripts mapped to up- and down-stream of introns relative to wild type. RNA-seq data from *osibm2_g2 #24* and WT (wild-type segregants of *osibm2*) were used. In each locus, the 5'/3' ratio of a representative transcript variant with TPM > 1 was used for calculation. Bars represent the means of DEGs with heterochromatic introns (n = 68), DEGs without heterochromatic intron (n = 335), and randomly selected 300 nonDEG loci \pm S.E.M. *p*-values were obtained by Tukey-Kramer test.
(PDF)

S13 Fig. Expression changes of genes in *osibm2*. (A) Representative rice genome loci showing altered expression patterns in mutants of *OsIBM2*. Tracks; Top to bottom: RNAseq (Reads per Million are indicated at top left), mCG ratio (0 to 1), mCHG ratio (0 to 1), mCHH ratio (0 to 1), H3K9me2 (RPM; 0 to 1), TE annotation (blue), repeats (orange), gene model (purple). The black arrow indicates the orientation of coding sequence. (B) Quantitative RT-PCR (qRT-PCR) analysis of expression of genes containing heterochromatic introns in *osibm2_g2#24* (*osibm2*) and WT (wild-type segregants of *osibm2*). Primer positions are indicated in Fig 5F and S13A Fig as red bars. Expression levels in each sample were normalized by *UBQ1* expression levels, and the average of *OsIBM2/UBQ1* in WT was set as 1. Bars represent the means of three biological replicates \pm S. D. (*n* = 3).
(PDF)

S14 Fig. 3' Rapid Amplification of cDNA Ends (RACE) of genes containing heterochromatin in mutants of *OsIBM2*. (A) 3' RACE of *Os01g0650200*. Upper panel: Structure of *Os01g0650200* locus and polyadenylated mRNA variants detected by 3' RACE. Exons and spliced introns confirmed by sequencing analysis are shown as black/red boxes and lines, respectively. Primer positions used for 3' RACE are indicated by arrowheads. Lower panel: Gel picture of DNA fragments amplified by 3' RACE. Two biological replicates for each genotype were examined. DNA fragments indicated by arrowheads were cloned and sequenced at least for 8 clones, and the representative sequences supported with more than 3 clones are shown in the upper panel. The black arrow indicates the orientation of coding sequence. NB: Nipponbare; *osibm2*: *osibm2_g2#24*; WT: wild-type segregants of *osibm2*; (A)n: polyadenylation. (B) 3' RACE of *Os06g0360600* as in (A). (C) 3' RACE of *Os08g0567200* as in (A). (D) The number of TEs showing expression changes in *osibm2_g2#24* (*osibm2*). 22 LTR TEs, and 1DNA/En-Spm showed significant changes (*q* < 0.05) of both up-regulation (12 TEs) and down-regulation (11 TEs). (E) Rice genome loci showing altered expression patterns of intronic TEs in mutants of *OsIBM2*. Tracks; Top to bottom: RNAseq (Reads per Million are indicated at top left), mCG ratio (0 to 1), mCHG ratio (0 to 1), mCHH ratio (0 to 1), H3K9me2 (RPM; 0 to 1), TE annotation (blue), repeats (orange), gene model (purple). The black arrow indicates the orientation of coding sequence.
(PDF)

S15 Fig. DNA methylation in *osibm2*. (A) Genome-wide DNA methylation in *osibm2_g2#24* (*osibm2*, T4) and their wild type segregating siblings (WT, T4) in CG, CHG and CHH contexts for each chromosome. Average methylation levels in 1 MB bins were plotted. (B) Metaplots of DNA methylation in *osibm2_g2#24* (*osibm2*) and their wild-type segregating siblings (WT) in CG, CHG and CHH contexts for indicated genome features.
(PDF)

S16 Fig. Rice homologs of the *Arabidopsis* H3K9 demethylase *IBM1*. Genome loci for *OsJMJ718* (*Os09g0393200*) (A) and *OsJMJ719* (*Os02g0109400*, *Os02G0109501*) (B). RNA-seq,

DNA methylation and H3K9me2 tracks are shown as in [S13 Fig. \(C\)](#) An alignment of amino acids sequences of *A. thaliana* IBM1 (At_IBM1) and OsJMJ718. The amino acid sequence of the N-terminal part of OsJMJ718 is predicted based on RNA-seq reads in this study. The alignment was generated by CLUSTAL W [131]. Jumonji-C (JmjC) domains predicted by SMART [132] are circled with blue lines. Positions of heterochromatic introns are indicated by red arrowheads.

(PDF)

S1 Table. Genes containing heterochromatic introns.

(XLSX)

S2 Table. Chromosomal positions of heterochromatic introns.

(XLSX)

S3 Table. Transposon annotation used in this study.

(XLSX)

S4 Table. Genes showing expression changes in *osibm2* mutants.

(XLSX)

S5 Table. Primers used in the study.

(XLSX)

S6 Table. A summary table for Whole Genome Bisulfite Sequencing (WGBS) analysis.

(XLSX)

S7 Table. A summary table for RNA-seq analysis.

(XLSX)

S1 Data. Numerical data used to generate Figures.

(XLSX)

Acknowledgments

We thank the Genebank project, NARO, for rice seed collection. We thank OIST SQC for BS-seq and RNA-seq analysis, and Drs. Yoshiki Habu, and Reina Komiya for critical reading of the manuscript. We also thank Dr. Steven D. Aird for editing the manuscript.

Author Contributions

Conceptualization: Nino A. Espinas, Hidetoshi Saze.

Data curation: Le Ngoc Tu, Shohei Takuno, Hidetoshi Saze.

Formal analysis: Nino A. Espinas, Le Ngoc Tu, Leonardo Furci, Yasuka Shimajiri, Yoshiko Harukawa, Saori Miura, Shohei Takuno.

Funding acquisition: Hidetoshi Saze.

Investigation: Nino A. Espinas, Le Ngoc Tu, Leonardo Furci, Yasuka Shimajiri, Yoshiko Harukawa, Saori Miura, Shohei Takuno.

Project administration: Hidetoshi Saze.

Supervision: Hidetoshi Saze.

Visualization: Le Ngoc Tu, Leonardo Furci, Shohei Takuno, Hidetoshi Saze.

Writing – original draft: Nino A. Espinas, Leonardo Furci, Shohei Takuno, Hidetoshi Saze.

Writing – review & editing: Nino A. Espinas, Le Ngoc Tu, Shohei Takuno, Hidetoshi Saze.

References

1. Kazazian HH Jr. Mobile elements: drivers of genome evolution. *Science*. 2004; 303(5664):1626–32. Epub 2004/03/16. <https://doi.org/10.1126/science.1089670> PMID: 15016989.
2. Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu Rev Plant Biol*. 2014; 65:505–30. Epub 2014/03/04. <https://doi.org/10.1146/annurev-arplant-050213-035811> PMID: 24579996.
3. Tenailon MI, Hollister JD, Gaut BS. A triptych of the evolution of plant transposable elements. *Trends Plant Sci*. 2010; 15(8):471–8. Epub 2010/06/15. <https://doi.org/10.1016/j.tplants.2010.05.003> PMID: 20541961.
4. Hollister JD, Gaut BS. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res*. 2009; 19(8):1419–28. Epub 2009/05/30. <https://doi.org/10.1101/gr.091678.109> PMID: 19478138; PubMed Central PMCID: PMC2720190.
5. Makarevitch I, Waters AJ, West PT, Stitzer M, Hirsch CN, Ross-Ibarra J, et al. Transposable elements contribute to activation of maize genes in response to abiotic stress. *PLoS Genet*. 2015; 11(1): e1004915. Epub 2015/01/09. <https://doi.org/10.1371/journal.pgen.1004915> PMID: 25569788; PubMed Central PMCID: PMC4287451.
6. Galindo-Gonzalez L, Mhiri C, Deyholos MK, Grandbastien MA. LTR-retrotransposons in plants: Engines of evolution. *Gene*. 2017; 626:14–25. Epub 2017/05/10. <https://doi.org/10.1016/j.gene.2017.04.051> PMID: 28476688.
7. Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, et al. Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. *Nature*. 2009; 461(7267):1130–4. Epub 2009/10/23. <https://doi.org/10.1038/nature08479> PMID: 19847266.
8. Lisch D. How important are transposons for plant evolution? *Nat Rev Genet*. 2013; 14(1):49–61. Epub 2012/12/19. <https://doi.org/10.1038/nrg3374> PMID: 23247435.
9. Quadrana L, Colot V. Plant Transgenerational Epigenetics. *Annu Rev Genet*. 2016; 50:467–91. Epub 2016/10/13. <https://doi.org/10.1146/annurev-genet-120215-035254> PMID: 27732791.
10. Slotkin RK, Martienssen R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet*. 2007; 8(4):272–85. Epub 2007/03/17. <https://doi.org/10.1038/nrg2072> PMID: 17363976.
11. Saze H, Mittelsten Scheid O, Paszkowski J. Maintenance of CpG methylation is essential for epigenetic inheritance during plant gametogenesis. *Nat Genet*. 2003; 34(1):65–9. Epub 2003/04/02. <https://doi.org/10.1038/ng1138> PMID: 12669067.
12. Hu LJ, Li N, Xu CM, Zhong SL, Lin XY, Yang JJ, et al. Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proc Natl Acad Sci USA*. 2014; 111(29):10642–7. <https://doi.org/10.1073/pnas.1410761111> WOS:000339310700060. PMID: 25002488
13. Kankel MW, Ramsey DE, Stokes TL, Flowers SK, Haag JR, Jeddelloh JA, et al. Arabidopsis MET1 cytosine methyltransferase mutants. *Genetics*. 2003; 163(3):1109–22. WOS:000182046900023. PMID: 12663548
14. Yamauchi T, Johzuka-Hisatomi Y, Terada R, Nakamura I, Iida S. The MET1b gene encoding a maintenance DNA methyltransferase is indispensable for normal development in rice. *Plant Mol Biol*. 2014; 85(3):219–32. <https://doi.org/10.1007/s11103-014-0178-9> WOS:000336030800002. PMID: 24535433
15. Matzke MA, Mosher RA. RNA-directed DNA methylation: an epigenetic pathway of increasing complexity. *Nat Rev Genet*. 2014; 15(6):394–408. Epub 2014/05/09. <https://doi.org/10.1038/nrg3683> PMID: 24805120.
16. Law JA, Jacobsen SE. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet*. 2010; 11(3):204–20. Epub 2010/02/10. <https://doi.org/10.1038/nrg2719> PMID: 20142834; PubMed Central PMCID: PMC3034103.
17. Wendte JM, Schmitz RJ. Specifications of Targeting Heterochromatin Modifications in Plants. *Mol Plant*. 2018; 11(3):381–7. <https://doi.org/10.1016/j.molp.2017.10.002> WOS:000426964100005. PMID: 29032247
18. Martienssen R, Moazed D. RNAi and heterochromatin assembly. *Cold Spring Harb Perspect Biol*. 2015; 7(8):a019323. Epub 2015/08/05. <https://doi.org/10.1101/cshperspect.a019323> PMID: 26238358; PubMed Central PMCID: PMC4526745.

19. Zemach A, Kim MY, Hsieh PH, Coleman-Derr D, Eshed-Williams L, Thao K, et al. The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell*. 2013; 153(1):193–205. Epub 2013/04/02. <https://doi.org/10.1016/j.cell.2013.02.033> PMID: 23540698; PubMed Central PMCID: PMC4035305.
20. Tan F, Zhou C, Zhou QW, Zhou SL, Yang WJ, Zhao Y, et al. Analysis of Chromatin Regulators Reveals Specific Features of Rice DNA Methylation Pathways. *Plant Physiol*. 2016; 171(3):2041–54. <https://doi.org/10.1104/pp.16.00393> WOS:000381303300043. PMID: 27208249
21. Numa H, Yamaguchi K, Shigenobu S, Habu Y. Gene Body CG and CHG Methylation and Suppression of Centromeric CHH Methylation are Mediated by DECREASE IN DNA METHYLATION1 in Rice. *Mol Plant*. 2015; 8(10):1560–2. Epub 2015/08/19. <https://doi.org/10.1016/j.molp.2015.08.002> PMID: 26277261.
22. Hirsch CD, Springer NM. Transposable element influences on gene expression in plants. *Biochim Biophys Acta*. 2017; 1860(1):157–65. Epub 2016/05/29. <https://doi.org/10.1016/j.bbagr.2016.05.010> PMID: 27235540.
23. Mirouze M, Vitte C. Transposable elements, a treasure trove to decipher epigenetic variation: insights from Arabidopsis and crop epigenomes. *J Exp Bot*. 2014; 65(10):2801–12. <https://doi.org/10.1093/jxb/eru120> WOS:000338005600021. PMID: 24744427
24. Henderson IR, Jacobsen SE. Tandem repeats upstream of the Arabidopsis endogene SDC recruit non-CG DNA methylation and initiate siRNA spreading. *Gene Dev*. 2008; 22(12):1597–606. <https://doi.org/10.1101/gad.1667808> WOS:000256797300006. PMID: 18559476
25. Soppe WJJ, Jacobsen SE, Alonso-Blanco C, Jackson JP, Kakutani T, Koornneef M, et al. The late flowering phenotype of *fwa* mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol Cell*. 2000; 6(4):791–802. [https://doi.org/10.1016/s1097-2765\(05\)00090-0](https://doi.org/10.1016/s1097-2765(05)00090-0) WOS:000090136700004. PMID: 11090618
26. Manning K, Tor M, Poole M, Hong Y, Thompson AJ, King GJ, et al. A naturally occurring epigenetic mutation in a gene encoding an SBP-box transcription factor inhibits tomato fruit ripening. *Nat Genet*. 2006; 38(8):948–52. <https://doi.org/10.1038/ng1841> WOS:000239325700027. PMID: 16832354
27. Gehring M, Bubb KL, Henikoff S. Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science*. 2009; 324(5933):1447–51. Epub 2009/06/13. <https://doi.org/10.1126/science.1171609> PMID: 19520961; PubMed Central PMCID: PMC2886585.
28. Le TN, Miyazaki Y, Takuno S, Saze H. Epigenetic regulation of intragenic transposable elements impacts gene transcription in Arabidopsis thaliana. *Nucleic Acids Res*. 2015; 43(8):3911–21. Epub 2015/03/31. <https://doi.org/10.1093/nar/gkv258> PMID: 25813042; PubMed Central PMCID: PMC4417168.
29. Seymour DK, Koenig D, Hagmann J, Becker C, Weigel D. Evolution of DNA methylation patterns in the Brassicaceae is driven by differences in genome organization. *PLoS Genet*. 2014; 10(11):e1004785. Epub 2014/11/14. <https://doi.org/10.1371/journal.pgen.1004785> PMID: 25393550; PubMed Central PMCID: PMC4230842.
30. Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, et al. The Norway spruce genome sequence and conifer genome evolution. *Nature*. 2013; 497(7451):579–84. Epub 2013/05/24. <https://doi.org/10.1038/nature12211> PMID: 23698360.
31. West PT, Li Q, Ji L, Eichten SR, Song J, Vaughn MW, et al. Genomic distribution of H3K9me2 and DNA methylation in a maize genome. *PLoS One*. 2014; 9(8):e105267. Epub 2014/08/15. <https://doi.org/10.1371/journal.pone.0105267> PMID: 25122127; PubMed Central PMCID: PMC4133378.
32. To TK, Saze H, Kakutani T. DNA Methylation within Transcribed Regions. *Plant Physiol*. 2015; 168(4):1219–25. Epub 2015/07/06. <https://doi.org/10.1104/pp.15.00543> PMID: 26143255; PubMed Central PMCID: PMC4528756.
33. Duan CG, Wang X, Zhang L, Xiong X, Zhang Z, Tang K, et al. A protein complex regulates RNA processing of intronic heterochromatin-containing genes in Arabidopsis. *Proc Natl Acad Sci U S A*. 2017; 114(35):E7377–E84. Epub 2017/08/16. <https://doi.org/10.1073/pnas.1710683114> PMID: 28808009; PubMed Central PMCID: PMC5584460.
34. Saze H, Kitayama J, Takashima K, Miura S, Harukawa Y, Ito T, et al. Mechanism for full-length RNA processing of Arabidopsis genes containing intragenic heterochromatin. *Nat Commun*. 2013; 4:2301. Epub 2013/08/13. <https://doi.org/10.1038/ncomms3301> PMID: 23934508.
35. Coustham V, Vlad D, Deremetz A, Gy I, Cubillos FA, Kerdaffrec E, et al. SHOOT GROWTH1 maintains Arabidopsis epigenomes by regulating IBM1. *PLoS One*. 2014; 9(1):e84687. Epub 2014/01/10. <https://doi.org/10.1371/journal.pone.0084687> PMID: 24404182; PubMed Central PMCID: PMC3880313.
36. Wang X, Duan CG, Tang K, Wang B, Zhang H, Lei M, et al. RNA-binding protein regulates plant DNA methylation by controlling mRNA processing at the intronic heterochromatin-containing gene IBM1.

- Proc Natl Acad Sci U S A. 2013; 110(38):15467–72. Epub 2013/09/05. <https://doi.org/10.1073/pnas.1315399110> PMID: 24003136; PubMed Central PMCID: PMC3780877.
37. Saze H. Epigenetic regulation of intragenic transposable elements: a two-edged sword. *J Biochem.* 2018; 164(5):323–8. <https://doi.org/10.1093/jb/mvy060> WOS:000449471000001. PMID: 30010918
 38. Wei L, Gu L, Song X, Cui X, Lu Z, Zhou M, et al. Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proc Natl Acad Sci U S A.* 2014; 111(10):3877–82. Epub 2014/02/21. <https://doi.org/10.1073/pnas.1318131111> PMID: 24554078; PubMed Central PMCID: PMC3956178.
 39. Liu N, Lee CH, Swigut T, Grow E, Gu B, Bassik MC, et al. Selective silencing of euchromatic L1s revealed by genome-wide screens for L1 regulators. *Nature.* 2018; 553(7687):228–32. Epub 2017/12/07. <https://doi.org/10.1038/nature25179> PMID: 29211708; PubMed Central PMCID: PMC5774979.
 40. Lorincz MC, Dickerson DR, Schmitt M, Groudine M. Intragenic DNA methylation alters chromatin structure and elongation efficiency in mammalian cells. *Nat Struct Mol Biol.* 2004; 11(11):1068–75. Epub 2004/10/07. <https://doi.org/10.1038/nsmb840> PMID: 15467727.
 41. Liu J, He Y, Amasino R, Chen X. siRNAs targeting an intronic transposon in the regulation of natural flowering behavior in *Arabidopsis*. *Genes Dev.* 2004; 18(23):2873–8. Epub 2004/11/17. <https://doi.org/10.1101/gad.1217304> PMID: 15545622; PubMed Central PMCID: PMC534648.
 42. Kum R, Tsukiyama T, Inagaki H, Saito H, Teraishi M, Okumoto Y, et al. The active miniature inverted-repeat transposable element mPing posttranscriptionally produces new transcriptional variants in the rice genome. *Mol Breeding.* 2015; 35(8). ARTN 159 <https://doi.org/10.1007/s11032-015-0353-y> WOS:000360005100010.
 43. Khan AR, Enjalbert J, Marsollier AC, Rousselet A, Goldringer I, Vitte C. Vernalization treatment induces site-specific DNA hypermethylation at the VERNALIZATION-A1 (VRN-A1) locus in hexaploid winter wheat. *BMC Plant Biol.* 2013; 13:209. Epub 2013/12/18. <https://doi.org/10.1186/1471-2229-13-209> PMID: 24330651; PubMed Central PMCID: PMC3890506.
 44. Osabe K, Harukawa Y, Miura S, Saze H. Epigenetic Regulation of Intronic Transgenes in *Arabidopsis*. *Sci Rep.* 2017; 7:45166. Epub 2017/03/25. <https://doi.org/10.1038/srep45166> PMID: 28338020; PubMed Central PMCID: PMC5364540.
 45. Tsuchiya T, Eulgem T. An alternative polyadenylation mechanism coopted to the *Arabidopsis* RPP7 gene through intronic retrotransposon domestication. *Proc Natl Acad Sci U S A.* 2013; 110(37):E3535–43. Epub 2013/08/14. <https://doi.org/10.1073/pnas.1312545110> PMID: 23940361; PubMed Central PMCID: PMC3773791.
 46. Ong-Abdullah M, Ordway JM, Jiang N, Ooi SE, Kok SY, Sarpan N, et al. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. *Nature.* 2015; 525(7570):533–7. Epub 2015/09/10. <https://doi.org/10.1038/nature15365> PMID: 26352475; PubMed Central PMCID: PMC4857894.
 47. Xie Y, Zhang Y, Han J, Luo J, Li G, Huang J, et al. The Intronic cis Element SE1 Recruits trans-Acting Repressor Complexes to Repress the Expression of ELONGATED UPPERMOST INTERNODE1 in Rice. *Mol Plant.* 2018; 11(5):720–35. Epub 2018/03/11. <https://doi.org/10.1016/j.molp.2018.03.001> PMID: 29524649.
 48. Questa JI, Song J, Geraldo N, An HL, Dean C. *Arabidopsis* transcriptional repressor VAL1 triggers Polycomb silencing at FLC during vernalization. *Science.* 2016; 353(6298):485–8. <https://doi.org/10.1126/science.aaf7354> WOS:000380583600040. PMID: 27471304
 49. Yuan WY, Luo X, Li ZC, Yang WN, Wang YZ, Liu R, et al. A cis cold memory element and a trans epigenome reader mediate Polycomb silencing of FLC by vernalization in *Arabidopsis*. *Nat Genet.* 2016; 48(12):1527–34. <https://doi.org/10.1038/ng.3712> WOS:000389011100013. PMID: 27819666
 50. Hong RL, Hamaguchi L, Busch MA, Weigel D. Regulatory elements of the floral homeotic gene AGAMOUS identified by phylogenetic footprinting and shadowing. *Plant Cell.* 2003; 15(6):1296–309. <https://doi.org/10.1105/tpc.009548> WOS:000185078300004. PMID: 12782724
 51. Song X, Cao X. Transposon-mediated epigenetic regulation contributes to phenotypic diversity and environmental adaptation in rice. *Curr Opin Plant Biol.* 2017; 36:111–8. Epub 2017/03/09. <https://doi.org/10.1016/j.pbi.2017.02.004> PMID: 28273484.
 52. Feschotte C, Jiang N, Wessler SR. Plant transposable elements: Where genetics meets genomics. *Nat Rev Genet.* 2002; 3(5):329–41. <https://doi.org/10.1038/nrg793> WOS:000175350000011. PMID: 11988759
 53. Matsumoto T, Wu JZ, Kanamori H, Katayose Y, Fujisawa M, Namiki N, et al. The map-based sequence of the rice genome. *Nature.* 2005; 436(7052):793–800. <https://doi.org/10.1038/nature03895> WOS:000231116500034. PMID: 16100779
 54. Du J, Zhong X, Bernatavichute YV, Stroud H, Feng S, Caro E, et al. Dual binding of chromomethylase domains to H3K9me2-containing nucleosomes directs DNA methylation in plants. *Cell.* 2012; 151

- (1):167–80. Epub 2012/10/02. <https://doi.org/10.1016/j.cell.2012.07.034> PMID: 23021223; PubMed Central PMCID: PMC3471781.
55. Roudier F, Ahmed I, Berard C, Sarazin A, Mary-Huard T, Cortijo S, et al. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. *EMBO J*. 2011; 30(10):1928–38. <https://doi.org/10.1038/emboj.2011.103> WOS:000291645400009. PMID: 21487388
 56. Bradnam KR, Korf I. Longer First Introns Are a General Property of Eukaryotic Gene Structure. *Plos One*. 2008; 3(8). ARTN e3093 <https://doi.org/10.1371/journal.pone.0003093> WOS:000264796800003. PMID: 18769727
 57. Oki N, Yano K, Okumoto Y, Tsukiyama T, Teraishi M, Tanisaka T. A genome-wide view of miniature inverted-repeat transposable elements (MITEs) in rice, *Oryza sativa ssp japonica*. *Genes Genet Syst*. 2008; 83(4):321–9. <https://doi.org/10.1266/ggs.83.321> WOS:000261872300004. PMID: 18931457
 58. Lu C, Chen J, Zhang Y, Hu Q, Su W, Kuang H. Miniature inverted-repeat transposable elements (MITEs) have been accumulated through amplification bursts and play important roles in gene expression and species diversity in *Oryza sativa*. *Mol Biol Evol*. 2012; 29(3):1005–17. Epub 2011/11/19. <https://doi.org/10.1093/molbev/msr282> PMID: 22096216; PubMed Central PMCID: PMC3278479.
 59. Quadrana L, Bortolini Silveira A, Mayhew GF, LeBlanc C, Martienssen RA, Jeddeloh JA, et al. The Arabidopsis thaliana mobilome and its impact at the species level. *Elife*. 2016; 5. Epub 2016/06/04. <https://doi.org/10.7554/eLife.15716> PMID: 27258693; PubMed Central PMCID: PMC4917339.
 60. Choi JY, Purugganan MD. Evolutionary Epigenomics of Retrotransposon-Mediated Methylation Spreading in Rice. *Mol Biol Evol*. 2018; 35(2):365–82. Epub 2017/11/11. <https://doi.org/10.1093/molbev/msx284> PMID: 29126199; PubMed Central PMCID: PMC5850837.
 61. Chen J, Hu Q, Zhang Y, Lu C, Kuang H. P-MITE: a database for plant miniature inverted-repeat transposable elements. *Nucleic Acids Res*. 2014; 42(Database issue):D1176–81. Epub 2013/11/01. <https://doi.org/10.1093/nar/gkt1000> PMID: 24174541; PubMed Central PMCID: PMC3964958.
 62. Sato Y, Takehisa H, Kamatsuki K, Minami H, Namiki N, Ikawa H, et al. RiceXPro Version 3.0: expanding the informatics resource for rice transcriptome. *Nuc Acids Res*. 2013; 41(D1):D1206–D13. <https://doi.org/10.1093/nar/gks1125> WOS:000312893300171. PMID: 23180765
 63. Kawahara Y, Oono Y, Wakimoto H, Ogata J, Kanamori H, Sasaki H, et al. TENOR: Database for Comprehensive mRNA-Seq Experiments in Rice. *Plant Cell Physiol*. 2016; 57(1):e7. Epub 2015/11/19. <https://doi.org/10.1093/pcc/pcv179> PMID: 26578693.
 64. Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biol*. 2005; 6(4). ARTN R33 <https://doi.org/10.1186/gb-2005-6-4-r33> WOS:000228436000010. PMID: 15833120
 65. Fawcett JA, Kado T, Sasaki E, Takuno S, Yoshida K, Sugino RP, et al. QTL map meets population genomics: an application to rice. *PLoS One*. 2013; 8(12):e83720. Epub 2014/01/01. <https://doi.org/10.1371/journal.pone.0083720> PMID: 24376738; PubMed Central PMCID: PMC3871663.
 66. Wasternack C, Hause B. Jasmonates: biosynthesis, perception, signal transduction and action in plant stress response, growth and development. An update to the 2007 review in *Annals of Botany*. *Ann Bot-London*. 2013; 111(6):1021–58. <https://doi.org/10.1093/aob/mct067> WOS:000319433300002. PMID: 23558912
 67. Mansueto L, Fuentes RR, Borja FN, Detras J, Abriol-Santos JM, Chebotarov D, et al. Rice SNP-seek database update: new SNPs, indels, and queries. *Nucleic Acids Res*. 2017; 45(D1):D1075–D81. Epub 2016/12/03. <https://doi.org/10.1093/nar/gkw1135> PMID: 27899667; PubMed Central PMCID: PMC5210592.
 68. Mei CS, Qi M, Sheng GY, Yang YN. Inducible overexpression of a rice allene oxide synthase gene increases the endogenous jasmonic acid level, PR gene expression, and host resistance to fungal infection. *Mol Plant Microbe In*. 2006; 19(10):1127–37. <https://doi.org/10.1094/Mpmi-19-1127> WOS:000240692300009. PMID: 17022177
 69. Ogawa S, Kawahara-Miki R, Miyamoto K, Yamane H, Nojiri H, Tsujii Y, et al. OsMYC2 mediates numerous defence-related transcriptional changes via jasmonic acid signalling in rice. *Biochem Bioph Res Co*. 2017; 486(3):796–803. <https://doi.org/10.1016/j.bbrc.2017.03.125> WOS:000399966700030. PMID: 28347822
 70. Le TN, Osabe K, Miyazaki Y, Saze H. Epigenetic regulation of intragenic repeats in plant genomes. *Genes Genet Syst*. 2016; 91(6):317–. WOS:000405886000006. <https://doi.org/10.1266/ggs.91.317> PMID: 28904303
 71. Yang Q, Liang C, Zhuang W, Li J, Deng H, Deng Q, et al. Characterization and identification of the candidate gene of rice thermo-sensitive genic male sterile gene *tms5* by mapping. *Planta*. 2007; 225(2):321–30. Epub 2006/08/10. <https://doi.org/10.1007/s00425-006-0353-6> PMID: 16896793.

72. Itabashi E, Iwata N, Fujii S, Kazama T, Toriyama K. The fertility restorer gene, *Rf2*, for Lead Rice-type cytoplasmic male sterility of rice encodes a mitochondrial glycine-rich protein. *Plant J.* 2011; 65(3):359–67. Epub 2011/01/27. <https://doi.org/10.1111/j.1365-313X.2010.04427.x> PMID: 21265890.
73. Kubo T, Takano-kai N, Yoshimura A. RFLP mapping of genes for long kernel and awn on chromosome 3 in rice. *Rice Genet Newsl.* 2001; 18:26–8.
74. Kang HG, Park S, Matsuoka M, An G. White-core endosperm floury endosperm-4 in rice is generated by knockout mutations in the C-type pyruvate orthophosphate dikinase gene (*OsPPDKB*). *Plant J.* 2005; 42(6):901–11. Epub 2005/06/09. <https://doi.org/10.1111/j.1365-313X.2005.02423.x> PMID: 15941402.
75. Hirano HY, Sano Y. Molecular Characterization of the Waxy Locus of Rice (*Oryza-Sativa*). *Plant Cell Physiol.* 1991; 32(7):989–97. <https://doi.org/10.1093/oxfordjournals.pcp.a078186> WOS: A1991GN75000009.
76. Kawakatsu T, Yamamoto MP, Touno SM, Yasuda H, Takaiwa F. Compensation and interaction between *RISBZ1* and *RPF* during grain filling in rice. *Plant Journal.* 2009; 59(6):908–20. <https://doi.org/10.1111/j.1365-313X.2009.03925.x> WOS:000269708400005. PMID: 19473328
77. Miura A, Nakamura M, Inagaki S, Kobayashi A, Saze H, Kakutani T. An Arabidopsis *jmjC* domain protein protects transcribed genes from DNA methylation at CHG sites. *EMBO J.* 2009; 28(8):1078–86. Epub 2009/03/06. <https://doi.org/10.1038/emboj.2009.59> PMID: 19262562; PubMed Central PMCID: PMC2653724.
78. Inagaki S, Miura-Kamio A, Nakamura Y, Lu F, Cui X, Cao X, et al. Autocatalytic differentiation of epigenetic modifications within the Arabidopsis genome. *EMBO J.* 2010; 29(20):3496–506. Epub 2010/09/14. <https://doi.org/10.1038/emboj.2010.227> PMID: 20834229; PubMed Central PMCID: PMC2964174.
79. Lu FL, Li GL, Cui X, Liu CY, Wang XJ, Cao XF. Comparative analysis of *JmjC* domain-containing proteins reveals the potential histone demethylases in Arabidopsis and rice. *J Integr Plant Biol.* 2008; 50(7):886–96. <https://doi.org/10.1111/j.1744-7909.2008.00692.x> WOS:000257708300014. PMID: 18713399
80. Zhang QJ, Zhu T, Xia EH, Shi C, Liu YL, Zhang Y, et al. Rapid diversification of five *Oryza AA* genomes associated with rice adaptation. *Proc Natl Acad Sci USA.* 2014; 111(46):E4954–E62. <https://doi.org/10.1073/pnas.1418307111> WOS:000345153300010. PMID: 25368197
81. Cheng C, Tarutani Y, Miyao A, Ito T, Yamazaki M, Sakai H, et al. Loss of function mutations in the rice chromomethylase *OsCMT3a* cause a burst of transposition. *Plant J.* 2015; 83(6):1069–81. Epub 2015/08/06. <https://doi.org/10.1111/tbj.12952> PMID: 26243209.
82. Moritoh S, Eun CH, Ono A, Asao H, Okano Y, Yamaguchi K, et al. Targeted disruption of an orthologue of DOMAINS REARRANGED METHYLASE 2, *OsDRM2*, impairs the growth of rice plants by abnormal DNA methylation. *Plant Journal.* 2012; 71(1):85–98. <https://doi.org/10.1111/j.1365-313X.2012.04974.x> WOS:000305407000008. PMID: 22380881
83. Higo H, Tahir M, Takashima K, Miura A, Watanabe K, Tagiri A, et al. *DDM1* (Decrease in DNA Methylation) genes in rice (*Oryza sativa*). *Molecular Genetics and Genomics.* 2012; 287(10):785–92. <https://doi.org/10.1007/s00438-012-0717-5> WOS:000309240500002. PMID: 22915302
84. Kakutani T, Jeddeloh JA, Flowers SK, Munakata K, Richards EJ. Developmental abnormalities and epimutations associated with DNA hypomethylation mutations. *Proc Natl Acad Sci USA.* 1996; 93(22):12406–11. <https://doi.org/10.1073/pnas.93.22.12406> WOS:A1996VP93700065. PMID: 8901594
85. Bartee L, Malagnac F, Bender J. Arabidopsis *cmt3* chromomethylase mutations block non-CG methylation and silencing of an endogenous gene. *Gene Dev.* 2001; 15(14):1753–8. <https://doi.org/10.1101/gad.905701> WOS:000170020000002. PMID: 11459824
86. Lindroth AM, Cao XF, Jackson JP, Zilberman D, McCallum CM, Henikoff S, et al. Requirement of *CHROMOMETHYLASE3* for maintenance of CpXpG methylation. *Science.* 2001; 292(5524):2077–80. <https://doi.org/10.1126/science.1059745> WOS:000169284700048. PMID: 11349138
87. Cao XF, Jacobsen SE. Role of the Arabidopsis DRM methyltransferases in de novo DNA methylation and gene silencing. *Curr Biol.* 2002; 12(13):1138–44. Pii S0960-9822(02)00925-9 [https://doi.org/10.1016/s0960-9822\(02\)00925-9](https://doi.org/10.1016/s0960-9822(02)00925-9) WOS:000176916900026. PMID: 12121623
88. Wei LY, Gu LF, Song XW, Cui XK, Lu ZK, Zhou M, et al. Dicer-like 3 produces transposable element-associated 24-nt siRNAs that control agricultural traits in rice. *Proc Natl Acad Sci USA.* 2014; 111(10):3877–82. <https://doi.org/10.1073/pnas.1318131111> WOS:000332564800056. PMID: 24554078
89. Niu XM, Xu YC, Li ZW, Bian YT, Hou XH, Chen JF, et al. Transposable elements drive rapid phenotypic variation in *Capsella rubella*. *Proc Natl Acad Sci USA.* 2019; 116(14):6908–13. <https://doi.org/10.1073/pnas.1811498116> WOS:000463069900067. PMID: 30877258

90. Maumus F, Quesneville H. Ancestral repeats have shaped epigenome and genome composition for millions of years in *Arabidopsis thaliana*. *Nat Commun*. 2014; 5. ARTN 4104 <https://doi.org/10.1038/ncomms5104> WOS:000338838200018. PMID: 24954583
91. Chamary JV, Hurst LD. Similar rates but different modes of sequence evolution in introns and at exonic silent sites in rodents: Evidence for selectively driven codon usage. *Mol Biol Evol*. 2004; 21(6):1014–23. <https://doi.org/10.1093/molbev/msh087> WOS:000221599300006. PMID: 15014158
92. Keightley PD, Gaffney DJ. Functional constraints and frequency of deleterious mutations in noncoding DNA of rodents. *Proc Natl Acad Sci USA*. 2003; 100(23):13402–6. <https://doi.org/10.1073/pnas.2233252100> WOS:000186573700053. PMID: 14597721
93. Parra G, Bradnam K, Rose AB, Korf I. Comparative and functional analysis of intron-mediated enhancement signals reveals conserved features among plants. *Nuc Acids Res*. 2011; 39(13):5328–37. <https://doi.org/10.1093/nar/gkr043> WOS:000293020000009. PMID: 21427088
94. Jeon JS, Lee S, Jung KH, Jun SH, Kim C, An G. Tissue-preferential expression of a rice alpha-tubulin gene, OsTubA1, mediated by the first intron. *Plant Physiol*. 2000; 123(3):1005–14. <https://doi.org/10.1104/pp.123.3.1005> WOS:000088213300023. PMID: 10889249
95. Morello L, Bardini M, Sala F, Breviaro D. A long leader intron of the Ostub16 rice beta-tubulin gene is required for high-level gene expression and can autonomously promote transcription both in vivo and in vitro. *Plant J*. 2002; 29(1):33–44. <https://doi.org/10.1046/j.0960-7412.2001.01192.x> WOS:000173544800004. PMID: 12060225
96. Liu SZ, Yeh CT, Ji TM, Ying K, Wu HY, Tang HM, et al. Mu Transposon Insertion Sites and Meiotic Recombination Events Co-Localize with Epigenetic Marks for Open Chromatin across the Maize Genome. *Plos Genet*. 2009; 5(11). ARTN e1000733 <https://doi.org/10.1371/journal.pgen.1000733> WOS:000272419500028. PMID: 19936291
97. Vollbrecht E, Duvick J, Schares JP, Ahern KR, Deewatthanawong P, Xu L, et al. Genome-Wide Distribution of Transposed Dissociation Elements in Maize. *Plant Cell*. 2010; 22(6):1667–85. <https://doi.org/10.1105/tpc.109.073452> WOS:000280505300004. PMID: 20581308
98. Yang L, Gaut BS. Factors that Contribute to Variation in Evolutionary Rate among *Arabidopsis* Genes. *Mol Biol Evol*. 2011; 28(8):2359–69. <https://doi.org/10.1093/molbev/msr058> WOS:000293304700017. PMID: 21389272
99. Turner BM. Epigenetic responses to environmental change and their evolutionary implications. *Philos T R Soc B*. 2009; 364(1534):3403–18. <https://doi.org/10.1098/rstb.2009.0125> WOS:000270800800009. PMID: 19833651
100. Meyers BC, Kaushik S, Nandety RS. Evolving disease resistance genes. *Curr Opin Plant Biol*. 2005; 8(2):129–34. Epub 2005/03/09. <https://doi.org/10.1016/j.pbi.2005.01.002> PMID: 15752991.
101. Espinas NA, Saze H, Saijo Y. Epigenetic Control of Defense Signaling and Priming in Plants. *Front Plant Sci*. 2016; 7. ARTN 1201 <https://doi.org/10.3389/fpls.2016.01201> WOS:000381206400001. PMID: 27563304
102. Hosaka A, Kakutani T. Transposable elements, genome evolution and transgenerational epigenetic variation. *Curr Opin Plant Biol*. 2018; 49:43–8. <https://doi.org/10.1016/j.gde.2018.02.012> WOS:000433211500007. PMID: 29525544
103. Li X, Guo K, Zhu X, Chen P, Li Y, Xie G, et al. Domestication of rice has reduced the occurrence of transposable elements within gene coding regions. *BMC Genomics*. 2017; 18(1):55. Epub 2017/01/11. <https://doi.org/10.1186/s12864-016-3454-z> PMID: 28068923; PubMed Central PMCID: PMC5223533.
104. Parenteau J, Maignon L, Berthoumieux M, Catala M, Gagnon V, Abou Elela S. Introns are mediators of cell response to starvation. *Nature*. 2019; 565(7741):612–7. Epub 2019/01/18. <https://doi.org/10.1038/s41586-018-0859-7> PMID: 30651641.
105. Morgan JT, Fink GR, Bartel DP. Excised linear introns regulate growth in yeast. *Nature*. 2019; 565(7741):606–11. Epub 2019/01/18. <https://doi.org/10.1038/s41586-018-0828-1> PMID: 30651636.
106. Sakai H, Lee SS, Tanaka T, Numa H, Kim J, Kawahara Y, et al. Rice Annotation Project Database (RAP-DB): An Integrative and Interactive Database for Rice Genomics. *Plant Cell Physiol*. 2013; 54(2):E6–+. <https://doi.org/10.1093/pcp/pcs183> WOS:000315218700006. PMID: 23299411
107. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res*. 2005; 110(1–4):462–7. <https://doi.org/10.1159/000084979> WOS:000231064600047. PMID: 16093699
108. Johnson M, Zaretskaya I, Raytselis Y, Merezuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res*. 2008; 36(Web Server issue):W5–9. Epub 2008/04/29. <https://doi.org/10.1093/nar/gkn201> PMID: 18440982; PubMed Central PMCID: PMC2447716.

109. Lu L, Chen JF, Robb SMC, Okumoto Y, Stajich JE, Wessler SR. Tracking the genome-wide outcomes of a transposable element burst over decades of amplification. *Proc Natl Acad Sci USA*. 2017; 114(49):E10550–E9. <https://doi.org/10.1073/pnas.1716459114> WOS:000417339700009. PMID: [29158416](https://pubmed.ncbi.nlm.nih.gov/29158416/)
110. Miki D, Shimamoto K. Simple RNAi vectors for stable and transient suppression of gene function in rice. *Plant Cell Physiol*. 2004; 45(4):490–5. <https://doi.org/10.1093/pcp/pch048> WOS:000221037200015. PMID: [15111724](https://pubmed.ncbi.nlm.nih.gov/15111724/)
111. Miura F, Enomoto Y, Dairiki R, Ito T. Amplification-free whole-genome bisulfite sequencing by post-bisulfite adaptor tagging. *Nucleic Acids Res*. 2012; 40(17):e136. Epub 2012/06/01. <https://doi.org/10.1093/nar/gks454> PMID: [22649061](https://pubmed.ncbi.nlm.nih.gov/22649061/); PubMed Central PMCID: [PMC3458524](https://pubmed.ncbi.nlm.nih.gov/PMC3458524/).
112. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30(15):2114–20. <https://doi.org/10.1093/bioinformatics/btu170> WOS:000340049100004. PMID: [24695404](https://pubmed.ncbi.nlm.nih.gov/24695404/)
113. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. 2011; 27(11):1571–2. <https://doi.org/10.1093/bioinformatics/btr167> WOS:000291062400018. PMID: [21493656](https://pubmed.ncbi.nlm.nih.gov/21493656/)
114. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010; 26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033> WOS:000275243500019. PMID: [20110278](https://pubmed.ncbi.nlm.nih.gov/20110278/)
115. Takuno S, Gaut BS. Body-methylated genes in *Arabidopsis thaliana* are functionally important and evolve slowly. *Mol Biol Evol*. 2012; 29(1):219–27. Epub 2011/08/05. <https://doi.org/10.1093/molbev/msr188> PMID: [21813466](https://pubmed.ncbi.nlm.nih.gov/21813466/).
116. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell*. 2013; 152(1–2):352–64. Epub 2013/01/15. <https://doi.org/10.1016/j.cell.2012.10.054> PMID: [23313553](https://pubmed.ncbi.nlm.nih.gov/23313553/); PubMed Central PMCID: [PMC3597350](https://pubmed.ncbi.nlm.nih.gov/PMC3597350/).
117. Ramirez F, Dundar F, Diehl S, Gruning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nuc Acids Res*. 2014; 42(W1):W187–W91. <https://doi.org/10.1093/nar/gku365> WOS:000339715000031. PMID: [24799436](https://pubmed.ncbi.nlm.nih.gov/24799436/)
118. Tian T, Liu Y, Yan HY, You Q, Yi X, Du Z, et al. agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nuc Acids Res*. 2017; 45(W1):W122–W9. <https://doi.org/10.1093/nar/gkx382> WOS:000404427000019. PMID: [28472432](https://pubmed.ncbi.nlm.nih.gov/28472432/)
119. Mi HY, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, et al. The PANTHER database of protein families, subfamilies, functions and pathways. *Nuc Acids Res*. 2005; 33:D284–D8. <https://doi.org/10.1093/nar/gki078> WOS:000226524300058. PMID: [15608197](https://pubmed.ncbi.nlm.nih.gov/15608197/)
120. Sun JQ, Nishiyama T, Shimizu K, Kadota K. TCC: an R package for comparing tag count data with robust normalization strategies. *BMC Bioinformatics*. 2013; 14. ArtN 219 <https://doi.org/10.1186/1471-2105-14-219> WOS:000321835900001. PMID: [23837715](https://pubmed.ncbi.nlm.nih.gov/23837715/)
121. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012; 9(4):357–U54. <https://doi.org/10.1038/nmeth.1923> WOS:000302218500017. PMID: [22388286](https://pubmed.ncbi.nlm.nih.gov/22388286/)
122. Alexandrov N, Tai S, Wang W, Mansueto L, Palis K, Fuentes RR, et al. SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res*. 2015; 43(Database issue):D1023–7. Epub 2014/11/29. <https://doi.org/10.1093/nar/gku1039> PMID: [25429973](https://pubmed.ncbi.nlm.nih.gov/25429973/); PubMed Central PMCID: [PMC4383887](https://pubmed.ncbi.nlm.nih.gov/PMC4383887/).
123. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015; 12(4):357–60. Epub 2015/03/10. <https://doi.org/10.1038/nmeth.3317> PMID: [25751142](https://pubmed.ncbi.nlm.nih.gov/25751142/); PubMed Central PMCID: [PMC4655817](https://pubmed.ncbi.nlm.nih.gov/PMC4655817/).
124. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009; 25(16):2078–9. Epub 2009/06/10. <https://doi.org/10.1093/bioinformatics/btp352> PMID: [19505943](https://pubmed.ncbi.nlm.nih.gov/19505943/); PubMed Central PMCID: [PMC2723002](https://pubmed.ncbi.nlm.nih.gov/PMC2723002/).
125. Freese NH, Norris DC, Loraine AE. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics*. 2016; 32(14):2089–95. Epub 2016/05/07. <https://doi.org/10.1093/bioinformatics/btw069> PMID: [27153568](https://pubmed.ncbi.nlm.nih.gov/27153568/); PubMed Central PMCID: [PMC4937187](https://pubmed.ncbi.nlm.nih.gov/PMC4937187/).
126. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nuc Acids Res*. 2013; 41(10). ARTN e108 <https://doi.org/10.1093/nar/gkt214> WOS:000319806600005. PMID: [23558742](https://pubmed.ncbi.nlm.nih.gov/23558742/)
127. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15(12). ARTN 550 <https://doi.org/10.1186/s13059-014-0550-8> WOS:000346609500022. PMID: [25516281](https://pubmed.ncbi.nlm.nih.gov/25516281/)

128. Gremme G, Brendel V, Sparks ME, Kurtz S. Engineering a software tool for gene structure prediction in higher organisms. *Inform Software Tech.* 2005; 47(15):965–78. <https://doi.org/10.1016/j.infsof.2005.09.005> WOS:000234322400003.
129. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and clustal X version 2.0. *Bioinformatics.* 2007; 23(21):2947–8. <https://doi.org/10.1093/bioinformatics/btm404> WOS:000251197700021. PMID: 17846036
130. Nei M, Gojobori T. Simple Methods for Estimating the Numbers of Synonymous and Nonsynonymous Nucleotide Substitutions. *Mol Biol Evol.* 1986; 3(5):418–26. WOS:A1986E136000004. <https://doi.org/10.1093/oxfordjournals.molbev.a040410> PMID: 3444411
131. Thompson JD, Higgins DG, Gibson TJ. Clustal-W—Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice. *Nuc Acids Res.* 1994; 22(22):4673–80. <https://doi.org/10.1093/nar/22.22.4673> WOS:A1994PU19900018. PMID: 7984417
132. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. *Nuc Acids Res.* 2018; 46(D1):D493–D6. <https://doi.org/10.1093/nar/gkx922> WOS:000419550700075. PMID: 29040681