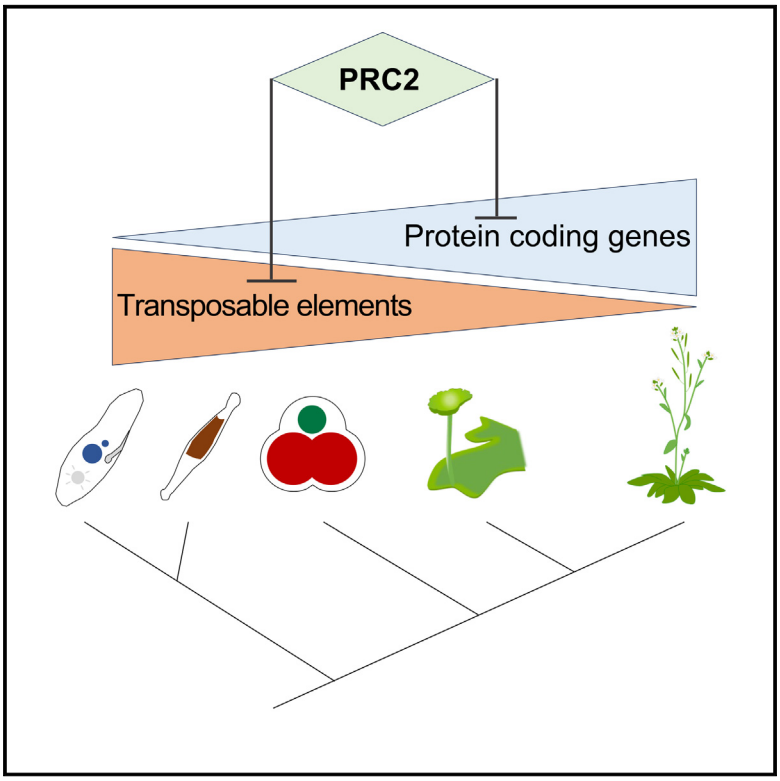
Article

Current Biology

The Polycomb repressive complex 2 deposits H3K27me3 and represses transposable elements in a broad range of eukaryotes

Graphical abstract



Authors

Tetsuya Hisanaga, Facundo Romani, Shuangyang Wu, ..., Leila Tirichine, Daniel Schubert, Frédéric Berger

Correspondence

danielschub@zedat.fu-berlin.de (D.S.), frederic.berger@gmi.oeaw.ac.at (F.B.)

In brief

The conserved Polycomb repressive complex 2 (PRC2) prevents transcription of protein-coding genes in eukaryotes. Hisanaga et al. show that PRC2 represses transposons in a broad range of eukaryotes, and its preference toward transposons over genes suggests that repressing transposons is an ancestral function.

Highlights

- The Polycomb repressive complex 2 represses TEs in a broad range of eukaryotes
- The proportion of TEs repressed by PRC2 decreases along the plant evolutionary tree
- H3K27me3-marked TEs in flowering plants contain transcription factor binding sites

Hisanaga et al., 2023, Current Biology *33*, 4367–4380
October 23, 2023
https://doi.org/10.1016/j.cub.2023.08.073

CellPress

Current Biology

CellPress
OPEN ACCESS

Article

The Polycomb repressive complex 2 deposits H3K27me3 and represses transposable elements in a broad range of eukaryotes

Tetsuya Hisanaga,[1] Facundo Romani,[2,11] Shuangyang Wu,[1,11] Teresa Kowar,[3,12] Yue Wu,[4,12] Ruth Lintermann,[3] Arie Fridrich,[1] Chung Hyun Cho,[1,5] Timothée Chaumier,[4] Bhagyshree Jamge,[1,6] Sean A. Montgomery,[1,6] Elin Axelsson,[1] Svetlana Akimcheva,[1] Tom Dierschke,[7] John L. Bowman,[7,8] Takayuki Fujiwara,[9,10] Shunsuke Hirooka,[9,10] Shin-ya Miyagishima,[9,10] Liam Dolan,[1] Leila Tirichine,[4] Daniel Schubert,[3,*] and Frédéric Berger[1,13,*]

[1]Gregor Mendel Institute, Austrian Academy of Sciences, Vienna BioCenter, Dr. Bohr-Gasse 3, 1030 Vienna, Austria
[2]Department of Plant Sciences, University of Cambridge, Cambridge CB2 3EA, UK
[3]Epigenetics of Plants, Institute of Biology, Freie Universität Berlin, 14195 Berlin, Germany
[4]Nantes Université, CNRS, US2B, UMR 6286, 44000 Nantes, France
[5]Department of Biological Sciences, Sungkyunkwan University, Suwon 16419, South Korea
[6]Vienna BioCenter PhD Program, Doctoral School of the University of Vienna and Medical University of Vienna, 1030 Vienna, Austria
[7]School of Biological Sciences, Monash University, Melbourne, VIC 3800, Australia
[8]ARC Centre of Excellence for Plant Success in Nature and Agriculture, Monash University, Clayton, Melbourne, VIC 3800, Australia
[9]Department of Gene Function and Phenomics, National Institute of Genetics, Mishima, Shizuoka 411-8540, Japan
[10]Department of Genetics, Graduate University for Advanced Studies, SOKENDAI, Mishima, Shizuoka 411-8540, Japan
[11]These authors contributed equally
[12]These authors contributed equally
[13]Lead contact
*Correspondence: danielschub@zedat.fu-berlin.de (D.S.), frederic.berger@gmi.oeaw.ac.at (F.B.)
https://doi.org/10.1016/j.cub.2023.08.073

SUMMARY

The mobility of transposable elements (TEs) contributes to evolution of genomes. Their uncontrolled activity causes genomic instability; therefore, expression of TEs is silenced by host genomes. TEs are marked with DNA and H3K9 methylation, which are associated with silencing in flowering plants, animals, and fungi. However, in distantly related groups of eukaryotes, TEs are marked by H3K27me3 deposited by the Polycomb repressive complex 2 (PRC2), an epigenetic mark associated with gene silencing in flowering plants and animals. The direct silencing of TEs by PRC2 has so far only been shown in one species of ciliates. To test if PRC2 silences TEs in a broader range of eukaryotes, we generated mutants with reduced PRC2 activity and analyzed the role of PRC2 in extant species along the lineage of Archaeplastida and in the diatom *P. tricornutum*. In this diatom and the red alga *C. merolae*, a greater proportion of TEs than genes were repressed by PRC2, whereas a greater proportion of genes than TEs were repressed by PRC2 in bryophytes. In flowering plants, TEs contained potential *cis*-elements recognized by transcription factors and associated with neighbor genes as transcriptional units repressed by PRC2. Thus, silencing of TEs by PRC2 is observed not only in Archaeplastida but also in diatoms and ciliates, suggesting that PRC2 deposited H3K27me3 to silence TEs in the last common ancestor of eukaryotes. We hypothesize that during the evolution of Archaeplastida, TE fragments marked with H3K27me3 were selected to shape transcriptional regulation, controlling networks of genes regulated by PRC2.

INTRODUCTION

A large fraction of eukaryotic genomes is composed of transposable elements (TEs) and sequences derived from ancient TEs that impact genome regulation and evolution.[1–4] The transcription of TE coding regions enables their movement (transposition). Because transposition may disrupt the function or the transcriptional regulation of protein-coding genes (PCGs), organisms tend to suppress TEs via the deposition of a specific type of chromatin that is enriched in DNA methylation (5-methylcytosine) and methylation of the lysine 9 of histone H3 (H3K9me).[5,6] In fission yeast, *Drosophila*, mammals, and flowering plants, H3K9me deposition is maintained by a positive feedforward loop involving small non-coding RNAs.[7–13] DNA methylation contributes to this feedback loop in mammals and flowering plants.[10,11,14] Eventually, H3K9me is bound by species-specific readers that prevent chromatin accessibility through some form of compaction[15,16] or phase separation,[17,18] thereby causing transcriptional silencing. In contrast to transcriptional silencing of TEs by H3K9 methylation, PCGs are temporarily silenced by the deposition of H2A ubiquitination by Polycomb repressive complex 1 (PRC1) and the deposition

of H3K27me3 by PRC2.[19–24] Recently, it was reported that PRC2 of the ciliate *Paramecium tetraurelia* is guided to TEs by a non-coding small RNA pathway and deposits both H3K9me3 and H3K27me3.[25,26] The presence of H3K27me3 on TEs was also reported in several species, including fungi,[27,28] the red alga *Cyanidioschyzon merolae*,[29,30] the model bryophyte *Marchantia polymorpha*,[31] and the diatom *Phaeodactylum tricornutum*,[32] suggesting that PRC2 might repress TEs in a broader range of eukaryotes.[33] However, whether PRC2 directly suppresses TEs in these species, which are positioned in distant branches across the eukaryotic tree of life, remains unknown. To answer this question, we used mutants with reduced PRC2 activity in *C. merolae*, *M. polymorpha*, and *P. tricornutum* and profiled the expression of TEs and PCGs. Here, we show that PRC2 actively represses the transcription of a higher proportion of TEs than PCGs in *P. tricornutum* and *C. merolae*. Although PRC2 deposits H3K27me3 on TEs in the two bryophytes *Anthoceros agrestis* and *M. polymorpha*, it primarily represses PCGs. In bryophytes and the flowering plant *Arabidopsis thaliana*, H3K27me3 covers TE fragments associated with PCGs, leading us to hypothesize the co-option of inactive TEs that became *cis*-elements with the capacity to bind specific transcription factors (TFs).

RESULTS

PRC2 represses TE expression in the diatom *P. tricornutum*

Previous genome-wide profiling of H3K27me3 in *P. tricornutum* showed the association of this mark not only with PCGs but also with a wide range of TEs and repeats.[32] To explore the role of H3K27me3 in transcriptional repression of TEs in *P. tricornutum*, we used a knockout (KO) mutant of the only ortholog of enhancer of zeste (E(z)), the catalytic subunit of PRC2.[32] We observed a similar proportion of overexpressed and repressed PCGs in Pt*e(z)* (Figure 1A) while TEs were primarily overexpressed (Figure 1B). We noted that the loss of Pt*E(z)* resulted in increased expression of 101 out of 132 TEs, and decreased expression of only 31 TEs (Figure 1C). About 12% (220) of PCGs with increased expression in Pt*e(z)* were covered by H3K27me3 in wild type (WT) (Figure 1C), indicating that PRC2 represses expression of a small proportion of PCGs via deposition of H3K27me3. By contrast, most (~94%) of all TEs with increased expression in Pt*e(z)* were covered by H3K27me3 in the WT (Figures 1C and 1D), supporting the hypothesis that PRC2 represses the expression of TEs. We defined PCGs and TEs that exhibited both increased expression in Pt*e(z)* and were covered by H3K27me3 in the WT as direct targets of PRC2 (159 PCGs and 95 TEs). Compared with the distribution of the different families of TEs, direct targets of PRC2 were mostly long terminal repeat (LTR) retrotransposons (Figure 1E). Both direct and indirect targets of H3K27me3 were scattered and evenly distributed along chromosomes (Figure 1F). Similarly, there was no apparent bias in the chromosomal localization of H3K27me3 for PCGs (Figure 1F). We conclude that in *P. tricornutum*, PRC2 deposits H3K27me3 evenly over chromosomes (with no discernable enriched domain) and represses expression of a higher proportion of TEs than PCGs.

PRC2 represses TE expression in the red alga *C. merolae*

To investigate the evolution of PRC2 function in the Archaeplastida lineage, we first selected the red alga *C. merolae*, which belongs to the class Cyanidiophyceae and which diverged from the Viridiplantae ca. 1.200 mya.[34] Previous genome-wide profiling of H3K27me3 demonstrated that this mark is associated not only with PCGs but also with TEs and repeats.[29] To re-analyze these data, we updated the TE annotation in the *C. merolae* genome (see details in STAR Methods). Half of the 5,074 TEs we annotated are class I retrotransposons, with 4.6% from the LINE family, 3.7% unclassified LTR, 30% Copia LTR, and 16% Gypsy LTR. The other half belonged to class II DNA transposons with 19% from the Harbinger family, 12% TcMar-ISRm11, and 5.7% TcMar (Figure S1A; Data S1). Analyses of genomic H3K27me3 enrichment profiles showed that 12% and 56% of H3K27me3 peaks overlapped with PCG and TE annotations, respectively, resulting in 3.6% of PCGs and 31% of TEs covered by H3K27me3, respectively (Figures 2A and 2B). There was no significant enrichment of specific TE families among the TEs covered by H3K27me3 (Figure S1A). Because 70% of the TEs were not covered by H3K27me3, we investigated the presence of other repressive marks such as methylation of lysine 9 of histone H3 (H3K9me1) and DNA (5methyl-C [5mC]), which mark TEs in many eukaryotes, including flowering plants.[12,13] We re-analyzed a genome-wide 5mC profile of *C. merolae*,[35] and confirmed that 5mC levels of the nuclear genome are not higher than the background levels measured on chloroplast and mitochondrial DNA (Table S1). We detected H3K9me1 via immunoblotting and confirmed the presence of this mark in the chromatin of *C. merolae* by acidic extraction of histones followed by mass spectrometry (Figure S1B; Data S2), suggesting its possible involvement in TE silencing. However, the genome of *C. merolae* contains only five genes encoding putative SET domain histone methyltransferases, four with homology to H3K4 methyltransferases and one with homology to the PRC2 methyltransferase E(z). Thus, none has homology to KRYPTONITE/SU(VAR)3-9 HOMOLOG 4, which is responsible for H3K9 methylation in *A. thaliana*. As a result, we were unable to interrogate the role of H3K9me1 and therefore focused our work on the role of H3K27me3, which is more broadly associated with TEs than with PCGs in *C. merolae*.

To explore the role of H3K27me3 in transcriptional repression of TEs in *C. merolae*, we disrupted the only ortholog of E(z), Cm*E(z)* (CMQ156C) (Figures S1C and S1D). Two independent loss-of-function alleles, Cm*e(z)-1* and Cm*e(z)-2*, exhibited a near complete loss of H3K27me3, with a concomitant reduction of H3K27me1 levels but no decreased levels of H3K9me1 (Figure 2C), suggesting that Cm*E(z)* deposits H3K27me3 but not H3K9me1. A comparison of transcriptomes of Cm*e(z)-1* and WT showed that expression levels of 0.4% (21) and 0.8% (42) PCGs out of a total 4,743 PCGs decreased and increased, respectively (q value < 0.05, $|\log_2$-fold change$| > 1$; Figures 2D–2F; Data S3). Two of 21 PCGs with decreased expression in Cm*e(z)-1* and about 60% (24) of PCGs with increased expression in Cm*e(z)-1* were covered by H3K27me3 in WT (Figure 2F), indicating that PRC2 represses expression of a small number of PCGs via deposition of H3K27me3. The loss of Cm*E(z)* also increased expression of 208 out of 5,074 TEs, whereas it resulted in decreased

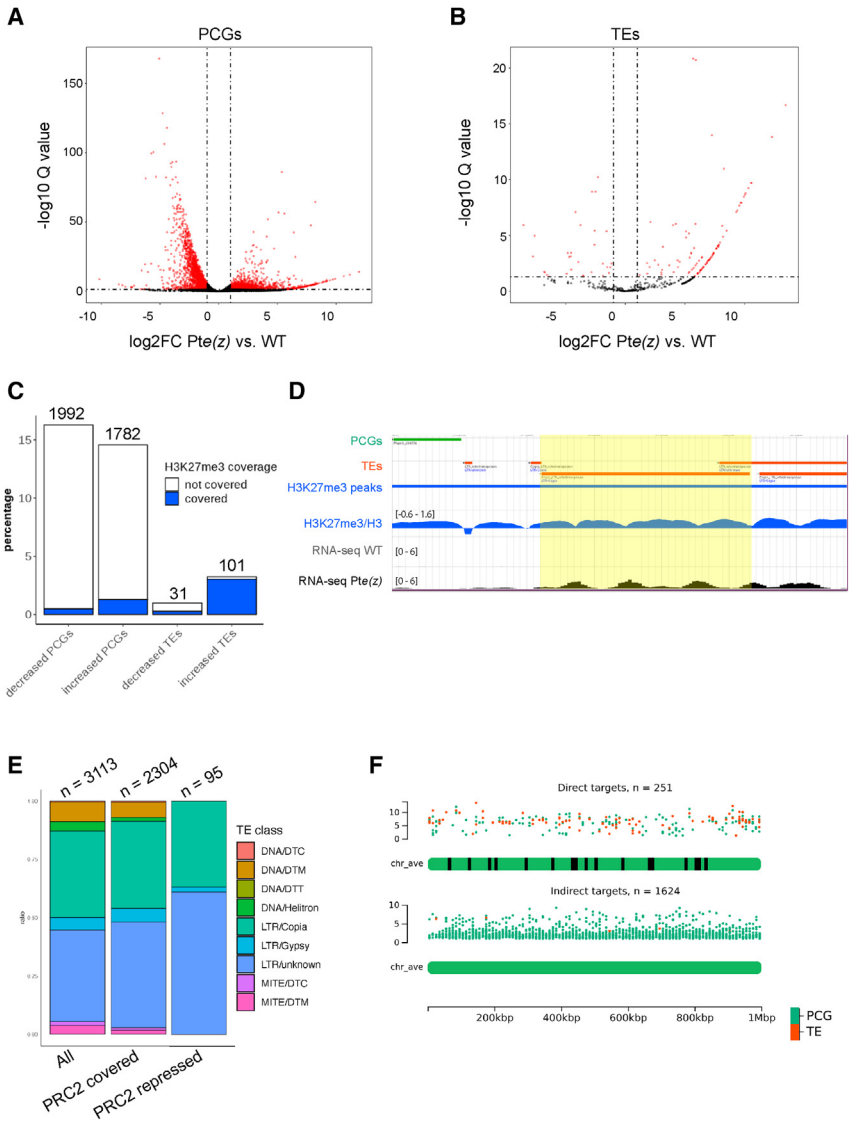


Figure 1. PRC2 primarily represses TE expression in the diatom *P. tricornutum*

(A) Volcano plot showing differential expression of protein-coding genes (PCGs) between wild-type and Pte*(z)* mutant. Differentially expressed PCGs (q value < 0.05 and $|\log_2$-fold change$| > 1$) are marked in red.

(B) Volcano plot showing differential expression of transposable elements (TEs) between wild-type and Pte*(z)* mutant. Differentially expressed TEs (q value < 0.05 and $|\log_2$-fold change$| > 1$) are marked in red.

(C) A bar plot indicating percentage of PCGs and TEs exhibiting decreased or increased expression levels in Pte*(z)* mutant in all PCGs and TEs. Those who are not covered or covered by H3K27me3 are shown in white or blue, respectively. Total numbers of PCGs or TEs in each category are shown on bars.

(D) A screenshot of the PhaeoEpiView browser displaying a representative region of H3K27me3 over TEs and genes in the wild type and Pte(z). The tracks show, from top to bottom, PCG annotation (green), TE annotation (orange), H3K27me3 peaks (blue), H3K27me3 coverage (blue, the y axis represents the logarithmic fold enrichment of H3K27me3, ranging from 1.6 to −0.6), expression level in wild type (gray), and Pte*(z)* mutant (black). The y axis for the RNA-seq track represents normalized transcripts per million.

(E) Stacked bar chart indicating proportion of TE families in all TE annotated (all TE), TEs covered by H3K27me3 (PRC2 covered), and TEs directly repressed by PRC2 (PRC2 repressed). Total numbers of TEs in each category are shown on bars.

(F) A chromosomal plot showing relative positions of PRC2 direct targets (top) and indirect targets (bottom). The x axis indicates relative positions of each target on an artificial chromosome that has an averaged size. The y axis indicates $\log_2$-fold change in RNA sequencing (RNA-seq) analysis comparing Pte*(z)* and wild type.

expression of only 30 TEs (Figures 2F and 2G; Data S3). More than half (115) of all TEs showing increased expression in Cme*(z)-1* were covered by H3K27me3 in the WT (Figure 2F), supporting the hypothesis that PRC2 represses the expression of TEs. We observed that the degree of derepression of TEs covered by H3K27me3 in the Cme*(z)* mutant (8%, 115/1,444) is similar to that in mutants deficient in H3K9 methylation in *A. thaliana* (Table 1).[36,37] The lack of expression of many TEs covered by H3K27me3 in absence of a H3K27me3 could be explained by various factors: other chromatin modifications may compensate for the absence of the silencing marks, or transcription may no longer be possible because the TE has lost essential *cis*-elements. We defined PCGs and TEs with higher expression in Cme*(z)* and covered by H3K27me3 in the WT as direct targets of PRC2 (24 PCGs and 115 TEs). We observed a slight enrichment of these likely direct PRC2 targets in the Gypsy family of retrotransposons ($p = 6.0 \times 10^{-4}$ in Fisher's exact test; Figure S1A). Most TEs repressed by PRC2 were located in subtelomeric chromosomal regions (Figure 2H) and it was thus possible that their repression by PRC2 was coincidental with a broader association of H3K27me3 with telomere function. By contrast, indirect targets of PRC2 that were misexpressed were scattered along chromosomes (Figure 2H). We conclude that PRC2 deposits H3K27me3 and represses expression of TEs preferentially in subtelomeric regions in *C. merolae*.

Association of H3K27me3 marks on TEs is conserved among bryophytes

In bryophytes, which diverged from vascular plants ca. 500–460 mya and comprise hornworts, liverworts, and mosses,[38] a fraction of TEs are covered by H3K27me3 in the liverwort *M. polymorpha*,[31] whereas TEs in the model moss *Physcomitrium patens* are covered mostly by H3K9me2.[39] To test if the association of H3K27me3 with TEs is conserved among bryophytes, we used the model hornwort *A. agrestis*, which diverged from bryophyte ancestors before the divergence of liverworts from mosses.[40] We annotated TEs in *A. agrestis* Oxford strain and identified 88,959 TEs, including 1,155 intact TEs belonging

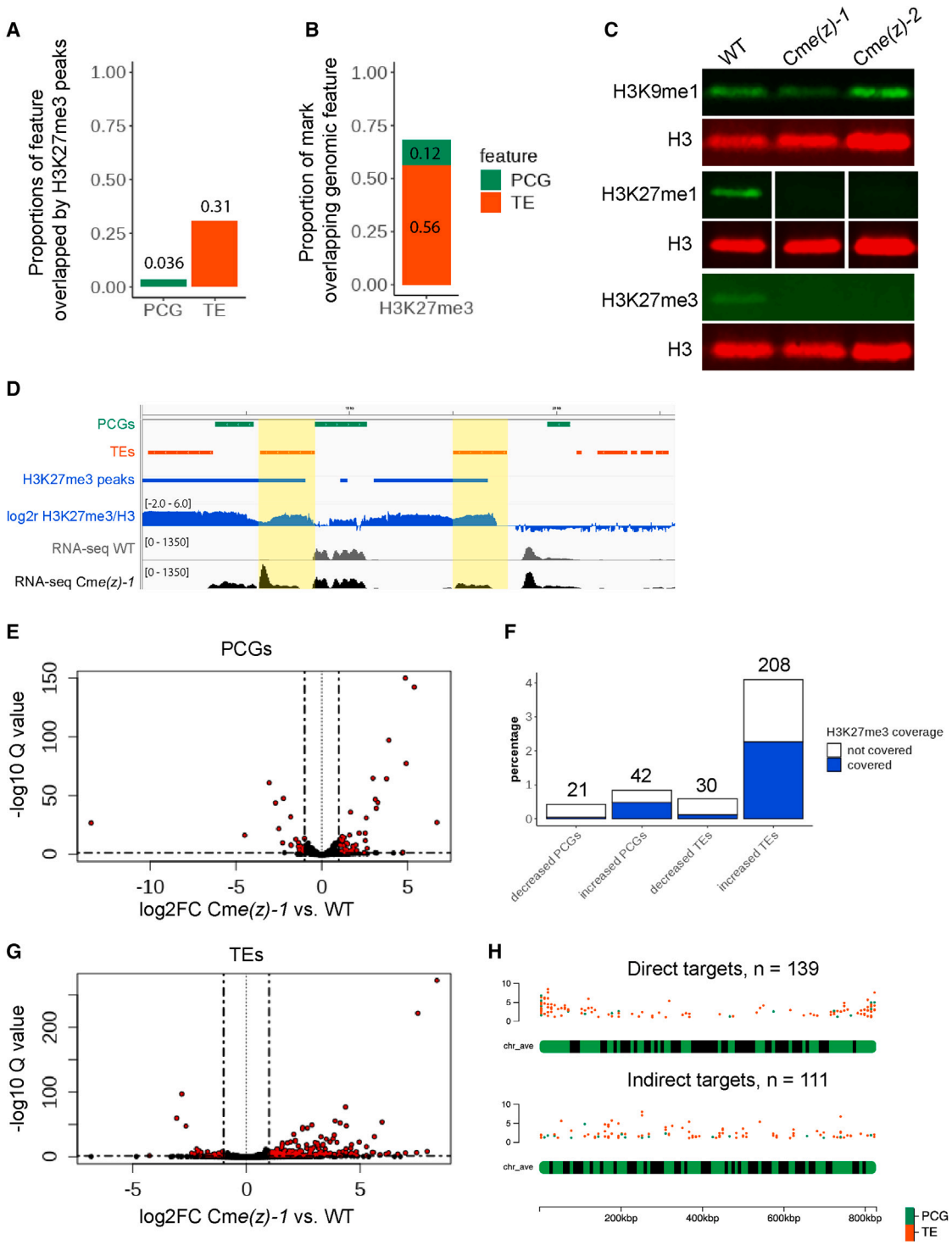


Figure 2. PRC2 primarily represses TE expression in the red alga

(A) Bar plot showing proportions of genomic features overlapped by H3K27me3 peaks.

(B) Bar plot showing proportions of H3K27me3 peaks overlapping to genomic features.

(C) Protein gel blot analyses indicating level of each histone modification in the wild type and two independent loss-of-function Cme*(z)* mutants.

(D) A screenshot of integrative genomics viewer showing subtelomeric region of chr 20. The tracks show, from top to bottom, PCG annotation (green), TE annotation (orange), H3K27me3 peaks (blue), H3K27me3 coverage (blue, the y axis represents the logarithmic fold enrichment of H3K27me3, ranging from 6 to −2; H3K27me3 coverage is normalized against the H3 ChIP signal), and expression levels in wild type (gray) and Cme*(z)-1* mutant (black). The y axis for the RNA-seq track represents read counts normalized by scaling factors from DEseq2.

(legend continued on next page)

to various TE families (Figure S1E; Data S1). Using chromatin immunoprecipitation coupled with DNA sequencing (ChIP-seq), we obtained genomic profiles of five post-translational histone modifications (PTMs) (H3K4me3, H3K36me3, H3K9me1, H3K27me1, and H3K27me3) and H3 from 4-week-old vegetative tissue of *A. agrestis* (see Hisanaga et al.[41] for a general overview of the chromatin of *A. agrestis*). We performed k-means clustering of chromatin marks over TEs and defined eight major TE clusters (Aa T1 to Aa T8) showing different chromatin environments (Figure 3A; Data S4). Clusters Aa T3, T4, and T5 contained 23%, 5.7%, and 5.5% of all TEs and were covered with H3K9me1 and H3K27me1 (Figure 3A). Clusters Aa T1 and Aa T2 comprised 5.9% and 6.3% of all TEs and were covered with either H3K4me3 and H3K27me3 (Aa T1) or H3K9me1 and H3K27me3 (Aa T2). We found that TEs from Aa T1 were enriched in unclassified TEs, and those from Aa T2 were enriched in Gypsy and Copia retrotransposons.[41] More than 80% of TEs in cluster Aa T1 were located close to genes belonging to PCG cluster Aa P1, which were also covered with H3K27me3 (Figures 3B and 3E). To compare these observations with *M. polymorpha*, we re-annotated TEs in the *M. polymorpha* Tak1v5.1_r2 genome using the same method as for *A. agrestis* and identified a total of 89,262 TEs, with fewer unclassified TEs compared with the previous annotation[31] (Figure 4E; Data S1). Using chromatin profiles obtained in the previous study,[31] we performed k-means clustering of the same set of chromatin marks used to study the chromatin of TEs in *A. agrestis*. The clustering identified five major TE clusters (Mp T1 to Mp T5) with different chromatin environments (Figure 3C; Data S4). As seen in *A. agrestis*, we observed a strong association of TEs with PCGs covered by a chromatin landscape (CL) dominated by H3K27me3 in *M. polymorpha* (5,719 TEs in cluster Mp T1 and cluster Mp P1; Figures 3C, 3D, and 3F[31]). Hence, the association of H3K27me3 with TEs is conserved in liverworts and hornworts. Because hornworts are sister to the other two groups that constitute the bryophyte lineage (mosses and liverworts),[40] we concluded that the association of TEs with H3K27me3 was ancestral in bryophytes.

PRC2 represses TE expression in *M. polymorpha*

To test whether PRC2 silences TEs in bryophytes, we used *M. polymorpha* as it is amenable to genetic manipulation.[42] To disrupt the function of PRC2, we focused on the ortholog of the PRC2 catalytic subunit E(z), which is encoded by three *E(z)* paralogs in *M. polymorpha*.[43] Only Mp*E(z)1* (Mp5g18040) is expressed in vegetative gametophytic tissues of *M. polymorpha*.[44] A knockdown of Mp*E(z)1* showed reduced growth and necrosis of tissues that hindered further analysis.[45] We observed that genes encoding the homeodomain TFs MpKNOX2 (Mp7g05320) and MpBELL1 (Mp8g18310), which are not expressed in the vegetative tissues of the gametophyte, were covered by H3K27me3 (Figure S2A). Since these TFs are essential for life phase transitions in plants,[46–49] we hypothesized that their misexpression in the Mp*e(z)1* mutant could be responsible for the lethality observed in the knockdown of Mp*E(z)1*. Therefore, we disrupted Mp*KNOX2* (see STAR Methods for details) to obtain null Mp*knox2-1* alleles in the Tak-2 female WT strain. We generated two KO alleles of Mp*e(z)1* in this mutant background, Mp*knox2-1* Mp*e(z)1-1* and Mp*knox2-1* Mp*e(z)1-2* (Figures S2B and S2C). We also used a construct expressing two guide RNAs targeting Mp*KNOX2* and Mp*E(z)1* to obtain the combination of two additional alleles, Mp*knox2-2* Mp*e(z)1-3*, in the Tak-1 male WT strain (Figures S2B and S2C). Although Mp*knox2* single-mutant alleles exhibited no developmental defects during vegetative growth phase, Mp*knox2* Mp*e(z)1* double mutants exhibited slower thallus growth but with a morphology similar to the WT (Figures S2D and S2E). This is consistent with our hypothesis that misexpression of MpKNOX2 was responsible for the lethality observed in the Mp*e(z)1* null mutant.

Western blot analyses using isolated nuclei from 14-day-old thalli indicated that H3K27me3 was undetectable in Mp*knox2-1* Mp*e(z)1-2*, whereas the levels of H3K9me1 and H3K27me1 were not reduced in Mp*knox2-1* Mp*e(z)1-2* when compared with WT (Figure 4A). In addition, the levels of these three post transcriptional modifications did not change in Mp*knox2-1* (Figure 4A), indicating a specific and likely complete loss of H3K27me3 in the Mp*e(z)1* KO mutant.

To evaluate the impact of the loss of H3K27me3 marks on the expression of PCGs and TEs, we compared the transcriptomes of the Mp*knox2-1* single mutant and the Mp*knox2-1* Mp*e(z)1-1* or Mp*knox2-1* Mp*e(z)1-2* double-mutant alleles using total RNA isolated from 14-day-old vegetative tissue. In both alleles of the double-mutant Mp*knox2* Mp*e(z)1*, 1,360 PCGs showed increased expression compared with the single-mutant Mp*knox2-1* (Figures 4B, S3A, and S3C; Data S3), and 77% (1,054) of these genes were covered by H3K27me3 in WT thallus (Figure S3E), supporting the hypothesis that PRC2 represses transcription of PCGs via deposition of H3K27me3. PCGs repressed by PRC2 encoded proteins involved in secondary metabolism and response to various stresses (Figure S3F), suggesting a role of PRC2 in the response of vegetative tissues to the environment. We also observed an impact of the loss of PRC2 on the expression of TEs, using a new set of annotated TEs in *M. polymorpha* (see details in STAR Methods). Expression of 506 TEs increased in both alleles of the double-mutant Mp*knox2* Mp*e(z)1* compared with the single-mutant Mp*knox2* (q value < 0.05, $\log_2$FC > 1; Figures 4C, S3B, and S3D; Data S3). We confirmed the increased expression of several TEs using quantitative real-time PCR analysis (Figure S3G). We observed

(E) Volcano plot showing differential expression of protein-coding genes (PCGs) between wild type and Cm*e(z)-1* mutant. Differentially expressed PCGs (q value < 0.05 and |$\log_2$-fold change| > 1) are marked in red.
(F) A bar plot indicating percentage of PCGs and TEs exhibiting decreased or increased expression levels in Cm*e(z)-1* mutant in all PCGs and TEs. Those who are not covered or covered by H3K27me3 are shown in white or blue, respectively. Total numbers of PCGs or TEs in each category are shown on bars.
(G) Volcano plot showing differential expression of transposable elements (TEs) between wild type and Cm*e(z)-1* mutant. Differentially expressed TEs (q value < 0.05 and |$\log_2$-fold change| > 1) are marked in red. See also Data S4.
(H) A chromosomal plot showing relative positions of PRC2 direct targets (top) and indirect targets (bottom). The x axis indicates relative positions of each target on an artificial chromosome that has an averaged size. The y axis indicates $\log_2$-fold change in RNA-seq analysis comparing Cm*e(z)-1* and wild type.
See also Figure S1, Table S1, and Data S1, S2, and S3.

Table 1. Proportion of TE fragments and TE genes per cluster showing increased expression in known silencing mutants of *A. thaliana*

Genotype	TE1 (%)	TE2 (%)	TE3 (%)	TE4 (%)	TE5 (%)	TE6 (%)	TE7 (%)	TEG1 (%)	TEG2 (%)	TEG3 (%)
ddm1	36.3	22.2	8.7	0.6	5.6	0.2	0.1	31.6	46.3	10.8
met1	26.2	14.9	7.0	0.7	5.5	0	0.1	21.7	35.5	9.4
suvh4/5/6	7.3	9.3	2.7	0.3	3.7	0.1	0.2	9.2	7.3	6.3

that 302 of the 506 TEs exhibiting increased expression in Mp*e(z)1* were primarily marked by H3K27me3 (cluster Mp T1, Figures 3C and 4D), suggesting that indirect effects caused the expression of 204 TEs in Mp*e(z)1*. If these 302 TEs were directly silenced by PRC2, we expected that they would not be expressed in the absence of the other major silencing pathway driven by DNA methylation. We tested this hypothesis by comparing the impact of the loss of PRC2 repression in Mp*e(z)1* with the impact of the loss of the main DNA methyltransferase in the mutant Mp*met*.[50] We confirmed that the loss of MpMET function caused overexpression of a higher proportion of TEs than PCGs (Figures S4A and S4B; Data S3). These TEs primarily belonged to the TE clusters Mp T2 and Mp T3, which are marked by H3K9me1 and 5mC, although some of them belonged to other clusters (Figures 3C, S4C, and S4D). Thus, this distribution was less specific than that observed in TEs overexpressed when H3K27me3 levels were reduced (Figure 4D). The overlap between TEs repressed by PRC2 and MET was remarkably small (27%), and it was further reduced when we focused on TEs from cluster Mp T1 (17%; Figures S4E and S4F). Hence, PRC2 directly and specifically represses a set of TEs independently from MpMET. We also noticed that Mp T2 comprised a large group of TEs marked by DNA methylation, H3K9me1, and H3K27me3 (Figures 3C and S4D). The redundancy between these three repressive epigenetic marks might explain why only very few TEs from this cluster are expressed in Mp*e(z)1* (Figure 4D).

The TEs repressed by PRC2 belonged to mutator DNA transposons and other uncategorized TE families (Figure 4E; $p = 3.5 \times 10^{-7}$ for mutator DNA transposons and $p < 2.2 \times 10^{-16}$ for uncategorized TE families in Fisher's exact test). Because TEs and PCGs are interspersed in the genome of *M. polymorpha*, we investigated whether PRC2 coregulated TEs with their closest neighboring PCGs. We observed that 89% of TEs from the cluster Mp T1 were located close to genes from the cluster Mp P1 (Figure 3F). Importantly, there was a significant enrichment of pairs of upregulated TEs and neighboring PCGs (Figure 4F). We conclude that, in *M. polymorpha*, PRC2 represses transcription from TEs, and these TEs are usually associated with a PCG that is also repressed by PRC2. It remains unclear if the recruitment of PRC2 by TEs causes silencing of neighboring PCGs or vice versa.

In *A. thaliana*, TEs covered by H3K27me3 contain TF-binding sites

In flowering plants, TEs are primarily marked by H3K9me1/2, However, in mutants lacking DNA methylation, H3K27me3 becomes associated with TEs, suggesting that a link between TEs and PRC2 is masked by the presence of marks of constitutive heterochromatin.[33] Based on a comprehensive nanalysis of chromatin states in *A. thaliana* seedlings, we established that 11% of TEs are covered by H3K27me3 (Figures S5A and S5B; Data S5, cluster TE6; clustering details in Jamge et al.[51]). In contrast to TEs from other clusters 1–4, which are covered with H3K9me1 and H3K27me1, these TEs are short and belong primarily to DNA and rolling circle (RC) transposon families (Figures S5C and S5D). Most of the TEs from cluster 6 were located close to PCGs covered by CL1, which is defined by a predominant enrichment of H3K27me3 as well as H2A.Z and H2AK121Ub[51] (Figures 5A and S5E). This observation suggested the hypothesis that these TEs might have been co-opted as *cis*-regulatory motifs for nearby PCGs. To test if these TEs could function as *cis*-regulatory elements controlling the repression of contiguous PCGs, we examined TF binding across all TEs in *A. thaliana* and calculated the enrichment of TF binding in vegetative or reproductive tissues from publicly available ChIP-seq experiments in each TE cluster (Figure 5B). In the TEs from cluster 6, which are marked by H3K27me3 and located close to PCGs, we observed a higher occupancy of TF binding than in the TE clusters 1–4, which are occupied by heterochromatin (Figure 5B, white). We thus hypothesized that H3K27me3-marked TEs might function as *cis*-regulatory elements. Moreover, this association was even clearer when we considered TF-binding events associated with the co-expression between the TF and the TE neighbor genes (Figure 5B, blue). By contrast, TF-binding events were not found in TEs from clusters 1–4, which are covered with constitutive heterochromatin (H3K9me2 and H3K27me1; Figure S5A). Altogether, these observations suggest that TFs associated with TEs are important for the regulation of neighboring genes. Among TFs expressed in flower buds and enriched in TEs from cluster 6, we found four MADS-box containing TFs (APETALA1, SEPALLATA3, AGAMOUS-LIKE15, and PISTILLATA) that control flower development (Figure 5C, top cluster).[52–56] A high proportion of TEs marked with H3K27me3 were located near PRC2-regulated PCGs covered by CL1, defined by enrichment in H2A.Z, H3K27me3, and H2AK121Ub, and excluded from actively transcribed genes associated with CLs 7–10 (defined in Jamge et al.[51]) (Figures 5A and 5D). We thus hypothesize that TEs harboring H3K27me3 contain *cis*-elements bound by TFs involved in the regulation of neighboring genes by PRC2.

We also observed a strong enrichment of TF binding in TEs from clusters 5 and 7 (Figure 5B). TEs from cluster 5 are short and marked by a hybrid chromatin state enriched in both H3K9me2 and H3K27me3 (Figures S5A and S5C). They are bound by TFs including the MADS-box factor APETALA3, the basic-helix-loop-helix (bHLH) factor GLABRA3, and the pioneer factor LEAFY[57,58] which are regulated by H3K27me3 and also control genes regulated by PRC2 and involved in flower development[52,59] (Figure 5C, bottom cluster).

The largest group of TFs enriched in TEs associated with TE cluster 7 (Figure 5B), which are short DNA and RC TEs, devoid of heterochromatin marks but with high accessibility at their

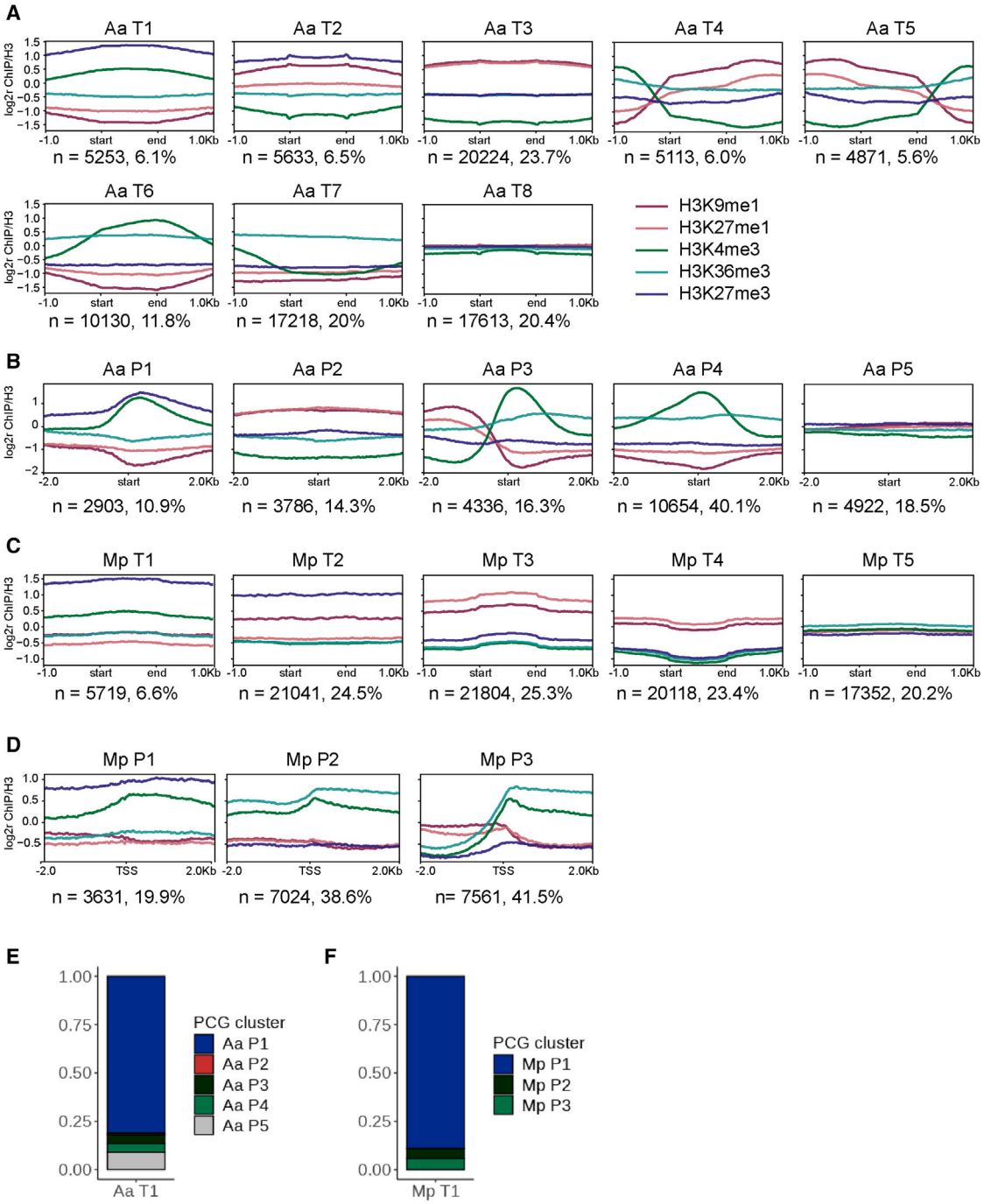


Figure 3. Associations of chromatin modifications of bryophytes

(A and B) Aggregate profile plots showing $\log_2$ ChIP/H3 enrichment for various chromatin modifications per TE (A) or PCG (B) cluster in *A. agrestis*. Total numbers of PCGs or TEs in each category are shown under plots.

(C and D) Aggregate profile plots showing $\log_2$ ChIP/H3 enrichment for various chromatin modifications per TE (C) or PCG (D) cluster in *M. polymorpha*. Total numbers of PCGs or TEs in each category are shown under plots.

(E and F) Stacked bar chart showing proportion of PCG clusters of nearby PCGs of TEs in clusters Aa T1 in *A. agrestis* (E) and Mp T1 in *M. polymorpha* (F). The nearest PCGs to each TE in the clusters Aa T1 and Mp T1 were identified and classified based on PCG clusters defined in (B) and (D), respectively.

See also Figure S1 and Data S1 and S4.

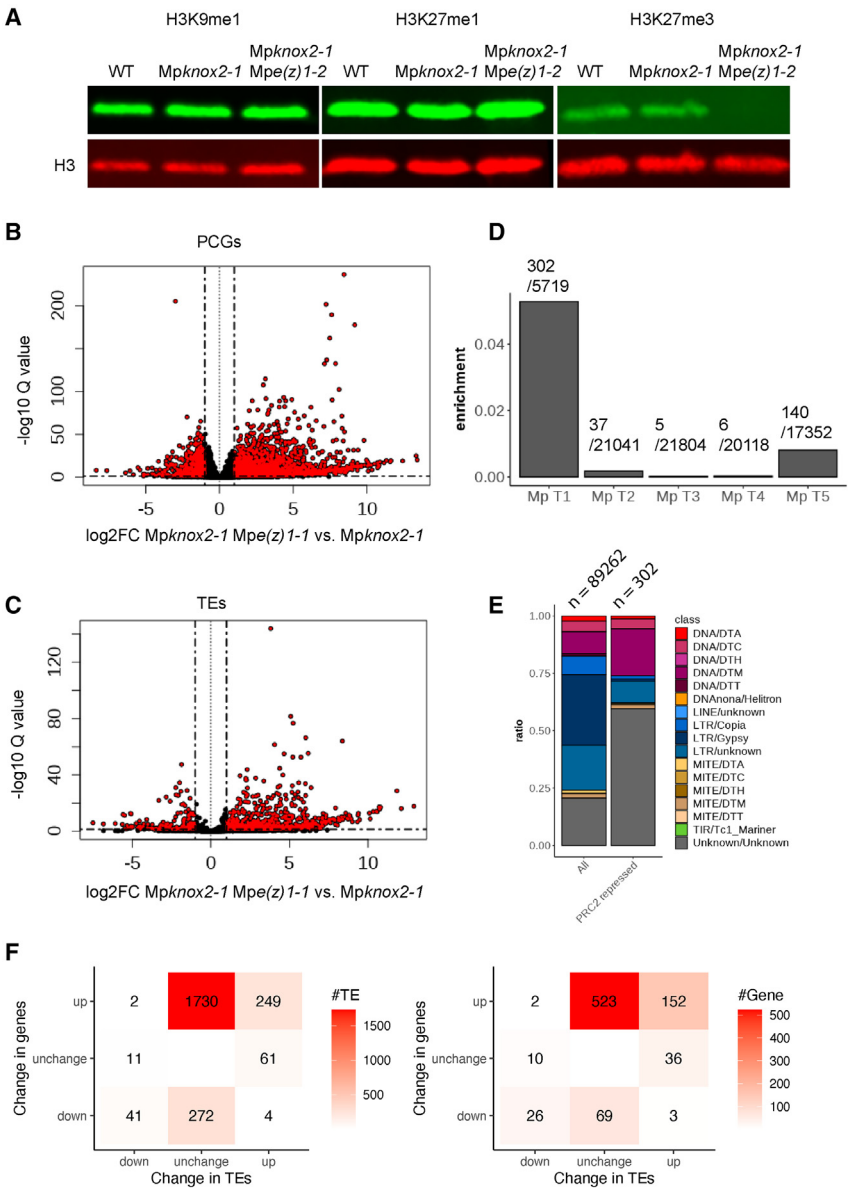


Figure 4. PRC2 silences some TEs in *M. polymorpha*

(A) Protein gel blot analyses indicating level of each histone modification in the wild type, Mp*knox2-1*, and Mp*knox2-1* Mp*e(z)1-2* mutant.

(B) Volcano plot of all protein-coding genes (PCGs) differentially expressed in Mp*knox2-1* Mp*e(z)1-1* compared with Mp*knox2-1*. Differentially expressed PCGs (q value < 0.05 and $|\log_2 FC| > 1$) are marked in red.

(C) Volcano plot of transposable elements (TEs) differentially expressed in Mp*knox2-1* Mp*e(z)1-1* compared with Mp*knox2-1*. Differentially expressed (q value < 0.05 and $|\log_2 FC| > 1$) TEs are marked in red.

(D) Bar plot indicating enrichment of upregulated TEs in each TE cluster, calculated as the number of TEs exhibiting increased expression in both Mp*knox2-1* Mp*e(z)1-1* and Mp*knox2-1* Mp*e(z)1-2* alleles in each cluster divided by the total number of TEs in each cluster.

(E) Stacked bar chart indicating proportion of TE families in all annotated TEs (all) and TEs directly repressed by PRC2 (PRC2 repressed). Total numbers of TEs in each category are shown on bars.

(F) The differential expression statistics of TE-PCG pairs covered by H3K27me3 in Mp*knox2-1* Mp*e(z)1-1*. Pairs of a TE and its nearest PCGs, both covered by the H3K27me3 mark, were selected based on Figure 3F. These PCGs and TEs were categorized by their expression status in the Mp*knox2-1* Mp*e(z)1-1* mutant compared with the wild type (up, down, or unchanged). This resulted in a total of nine combinations (3 states of genes × 3 states of TEs). The x axis represents the three states of TEs, and the y axis represents the three states of PCGs. The left panel displays the TE number, whereas the right panel displays the gene number. p value of the upregulated TEs and neighboring PCGs pair is 0 (hypergeometric test). p values for the other intersects can be found in Table S2.

See also Figures S2–S4, Table S2, and Data S1 and S3.

boundaries (Figures S5A, S5C, S5D, and S5F). These TEs were strongly associated with PCGs covered by either CL6 or CL8 (Figures 5D and S5E), suggesting that they also contain TF-binding *cis*-elements controlling expression of neighboring PCGs. Overall, our results lead us to hypothesize that a large number of TEs might have been domesticated in *A. thaliana*, providing *cis*-elements involved in activation (cluster TE7) or PRC2-targeted repression (clusters TE5 and TE6) of PCGs.

DISCUSSION

Here, we have shown that PRC2 silences TEs among diverse groups of Archaeplastida. Silencing might occur either via the extension of a region covered by H3K27me3 to TEs in its neighborhood or via an indirect impact on silencing from other epigenetic marks. However, in the red alga *C. merolae*, PRC2 deposits H3K27me3 and silences a subset of the TEs. Although the proportion of TEs covered by H3K27me3 in bryophytes and angiosperms is less prominent than in *C. merolae*, it is likely that the deposition of H3K27me3 on TEs by PRC2 is common in Archaeplastida (Figure 6). In addition, we show that PRC2 silences more TEs than PCGs in the diatom *P. cornutum*, which belongs to the SAR supergroup. Similarly, ciliates, also members of SAR, use PRC2 to silence more TEs than PCGs.[26] Among Opisthokonta, PRC2 silences a broad spectrum of TEs in female primordial germ cells,[60] and H3K27me3 is associated with TEs in fungi[27,28] and a rotifer.[61] PRC2 is present in most eukaryotes and is considered one of the most conserved epigenetic writers.[62–64] Although it is possible that PRC2 was independently recruited to silence TEs in four of the main groups of eukaryotes, the presence of TEs marked and repressed by PRC2 across four major clades of eukaryotes rather suggests that PRC2 targeted TEs for silencing in the last common ancestor of eukaryotes (Figure 6).

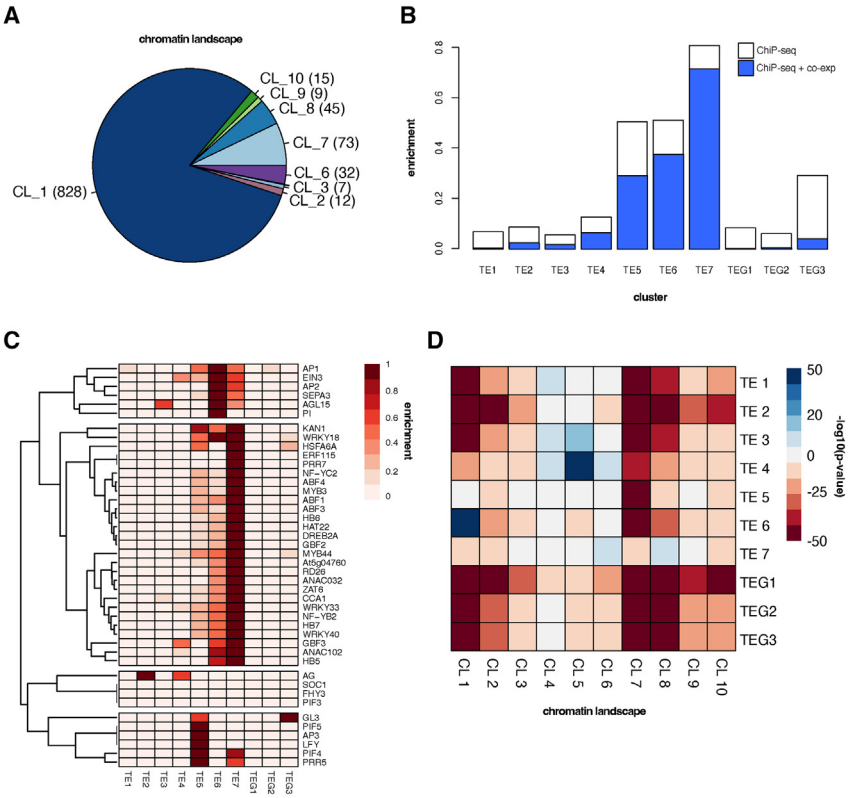


Figure 5. TE fragments covered by H3K27me3 are enriched in transcription factors binding sites in *A. thaliana*

(A) Pie chart showing proportion of chromatin landscapes of PCGs neighbors of TEs in the cluster TE6 in *A. thaliana*.

(B) Bar graph showing enrichment of TF-binding sites in each TE and TEG (transposon gene) cluster. White bars indicate TF binding based on all binding events analyzed, and blue bars indicate only binding events associated with co-expression of the downstream gene. See Jamge et al.[51] and the STAR Methods section for details of clustering analyses of TEs and TEGs in *A. thaliana*. Cluster assignments can be found in Data S5.

(C) Heatmap showing enrichment of every TF analyzed in each TE and TEG cluster.

(D) Heatmap showing statistical enrichment of every TE and TEG cluster in each chromatin landscape described in Jamge et al.[51]

See also Figure S5 and Data S5.

In the ciliate *P. tetraurelia*, recruitment of PRC2[26] involves non-coding RNAs[25] in a manner reminiscent of the recruitment of the machinery that deposits H3K9me3 in fungi and animals.[65] In the red alga *C. merolae*, H3K27me3-silenced TEs are associated with the telomeres, reminiscent of PRC2 recruitment by telomere repeat binding (TRB) factors at telobox motifs in angiosperms.[66] The mechanism by which PRC2 is recruited to induce silencing in *P. tricornutum* could not be identified based on orthologs of proteins recruiting PRC2 in other species. The intriguing observation of distinct patterns of H3K27me3 localization across chromosomes further supports the idea that diverse mechanisms for recruiting PRC2 have been selected to achieve silencing of TEs across eukaryotes.

It has been proposed that remnants of TEs, also called fossil TEs, became *cis*-elements bound by TFs during eukaryotic evolution.[67] This scenario has been validated by numerous examples in mammals and accompanied the evolution of mammalian-specific developmental features, such as the placenta and specific aspects of brain development.[3] In *A. thaliana*, a few examples of TE fragments have been associated with specific TF-binding sites,[68,69] and TEs have been shown to influence expression of nearby genes and mediate the emergence of adaptive traits.[70–73] The presence of TF-binding sites on several hundred short TEs present outside of constitutive heterochromatin supports the hypothesis that fossil TEs were domesticated during plant evolution, thus providing an important source of *cis*-elements for TFs that activate or repress transcription.

Several examples in plants have illustrated that TEs recruit DNA methylation and H3K9 methylation that affect the expression of nearby PCGs.[67,74,75] These events have been associated with selection of traits in crops.[76–79] Similarly, our results lead to the hypothesis that the recruitment of the ancestral transcriptional repressive Polycomb machinery by TEs has been co-opted to silence PCGs. The observed distribution pattern suggests that H3K27me3-mediated repression of TEs in *P. tricornutum* may have a functional role in repressing neighboring genes, although the exact mechanism behind this has yet to be elucidated. In bryophytes, H3K27me3 co-regulates numerous pairs of TEs and PCGs distributed uniformly across the chromosome. Hence, we hypothesize that TEs contain or behave as *cis*-elements controlling the transcriptional activities of PCGs. In *A. thaliana*, we show that TEs contain binding sites for TFs regulating flower development. In vegetative tissues, these TEs and the contiguous genes are repressed by PRC2. PRC2 is likely recruited by other TFs or transcriptional inhibitors, as shown in other studies.[80,81] It is also possible that the same factor recruits first PRC2 at an early developmental point and then recruits activators of transcription at a later point, as shown for TRB factors.[82] Interestingly, PRC2 controls the expression of the TFs controlling flower development with binding sites in TEs. Hence, we hypothesize that fossil TEs might have been the source of Polycomb gene networks that control multiple aspects of development and response to environmental cues in land plants.[39,83,84] The ancestral role of PRC2 in TE silencing in distant lineages of eukaryotes suggests that similar evolutionary trajectories might also be observed in phyla other than Archaeplastida.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability

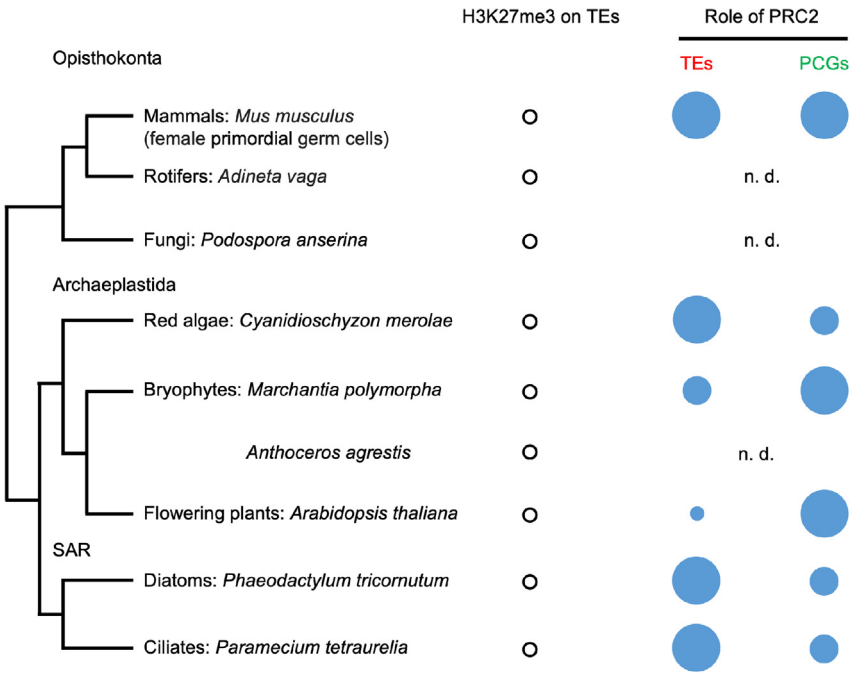


Figure 6. PRC2 represses TEs in various eukaryotes

A simplified phylogenetic relationship illustrates the evolutionary conservation of the role of PRC2 in TE silencing with considerable variation among eukaryotes. Depicted is the relative number of repressed TEs and PCGs, respectively, in the red alga *C. merolae*, the bryophyte *M. polymorpha*, the flowering plant *A. thaliana*, the diatom *P. tricornutum*, and the ciliate *P. tetraurelia*. The relative numbers of TEs and PCGs deregulated in mouse primordial germ cells were estimated by Huang et al.[60]

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cub.2023.08.073.

ACKNOWLEDGMENTS

This manuscript is dedicated to Dr. Harmit Malik, who hypothesized the ancestral association of PRC2 with TE silencing at the Gordon Epigenetic Conference in 2017. We thank Matt Watson for suggestions and critical reading of the manuscript. F.B. acknowledges support from the PlantS, Proteomics Facility (especially Otto Hudecz and Elisabeth Roitinger), Next Generation Sequencing and Histopathology facilities at the Vienna BioCenter Core Facilities (VBCF), and the BioOptics facility and Molecular Biology Services from the Institute for Molecular Pathology (IMP), members of the Vienna BioCenter (VBC), Austria. Proteomics analyses were performed by the Proteomics Facility using the VBCF instrument pool. This work was funded by FWF grants P32054 and P33380 to F.B., FWF doctoral school DK W1238 to S.A.M., the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014–2020) under the Marie Curie, Skłodowska Grant Agreement no. 847548 (VIP2) to T.H. and S.W., the Biotechnology and Biological Sciences Research Council BB/T007117/1 to F.R., the Australian Research Council DP210101423 to J.L.B., and a European Research Council Advanced Grant (project number 787613) from the European Commission to L.D. D.S. acknowledges support by CRC973 (DFG). L.T. acknowledges support from the region of Pays de la Loire (ConnecTalent EPIALG project) and Epicycle ANR (ANR-19-CE20-0028-02). Y.W. was supported by a PhD fellowship from the Chinese Scholarship Council (CSC-201908340073). C.H.C. was supported by the Basic Science Research Program through the National Research Foundation of Korea (RS-2023-00248097). T.K. was supported by an Elsa-Neumann-fellowship.

AUTHOR CONTRIBUTIONS

T.H., D.S., and F.B. conceived and designed the experiments. T.F., S.H., and S.-y.M. obtained the *Cme(z)-1* mutant in *C. merolae*. T.K. and R.L. obtained the *Cm(e)z-2* mutant. T.K. performed the RNA-seq of *Cme(z)-1*. R.L. generated western blots in *C. merolae* and *M. polymorpha*. A.F. and C.H.C. performed mass spectrometry analyses of histone modifications in *C. merolae*. T.H. and S.A.M. obtained the KO mutants in *M. polymorpha* with help from S.A. T.H. performed RNA-seq of *M. polymorpha*. A.F. performed whole-genome bisulfite-seq of *M. polymorpha*. F.R. analyzed the TFs present in TEs in *A. thaliana* marked by H3K27me3 selected by B.J. T.D. and J.L.B. obtained data that led to the strategy from KO PRC2 in *M. polymorpha*. S.W. annotated TEs in the genomes of *C. merolae*, *A. agrestis*, and *M. polymorpha*. T.H. and

S.W. performed the analysis of the data with help from E.A., who also performed statistical analyses and curated data. Y.W. performed the RNA-seq data in Pte*(z)* and WT. Y.W. and T.C. performed the related bioinformatic analysis. Y.W., T.C., and L.T. generated Figure 1. L.T. supervised Y.W. and T.C. L.T. edited the manuscript. F.B. and D.S. supervised the study. T.H. and F.B. wrote the manuscript draft. The draft was revised with input from D.S., J.L.B., and L.D.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: October 28, 2022
Revised: June 19, 2023
Accepted: August 24, 2023
Published: September 21, 2023

REFERENCES

1. Bourque, G., Burns, K.H., Gehring, M., Gorbunova, V., Seluanov, A., Hammell, M., Imbeault, M., Izsvák, Z., Levin, H.L., Macfarlan, T.S., et al. (2018). Ten things you should know about transposable elements. Genome Biol. *19*, 199.
2. Cosby, R.L., Chang, N.-C., and Feschotte, C. (2019). Host-transposon interactions: conflict, cooperation, and cooption. Genes Dev. *33*, 1098–1116.
3. Fueyo, R., Judd, J., Feschotte, C., and Wysocka, J. (2022). Roles of transposable elements in the regulation of mammalian transcription. Nat. Rev. Mol. Cell Biol. *23*, 481–497.
4. Baduel, P., and Colot, V. (2021). The epiallelic potential of transposable elements and its evolutionary significance in plants. Philos. Trans. R. Soc. Lond. B Biol. Sci. *376*, 20200123.
5. Almeida, M.V., Vernaz, G., Putman, A.L.K., and Miska, E.A. (2022). Taming transposable elements in vertebrates: from epigenetic silencing to domestication. Trends Genet. *38*, 529–553.
6. Janssen, A., Colmenares, S.U., and Karpen, G.H. (2018). Heterochromatin: guardian of the genome. Annu. Rev. Cell Dev. Biol. *34*, 265–288.
7. Martienssen, R., and Moazed, D. (2015). RNAi and heterochromatin assembly. Cold Spring Harb. Perspect. Biol. *7*, a019323.
8. Shipkovenska, G., Durango, A., Kalocsay, M., Gygi, S.P., and Moazed, D. (2020). A conserved RNA degradation complex required for spreading and epigenetic inheritance of heterochromatin. eLife *9*, e54341, https://doi.org/10.7554/eLife.54341.
9. Andersen, P.R., Tirian, L., Vunjak, M., and Brennecke, J. (2017). A heterochromatin-dependent transcription machinery drives piRNA expression. Nature *549*, 54–59.
10. Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220.
11. Stamidis, N., and Żylicz, J.J. (2023). RNA-mediated heterochromatin formation at repetitive elements in mammals. EMBO J. *42*, e111717.
12. Jamge, B., and Berger, F. (2022). Diversification of chromatin organization in eukaryotes. Curr. Opin. Cell Biol. *74*, 1–6.
13. Padeken, J., Methot, S.P., and Gasser, S.M. (2022). Establishment of H3K9-methylated heterochromatin and its functions in tissue differentiation and maintenance. Nat. Rev. Mol. Cell Biol. *23*, 623–640. https://doi.org/10.1038/s41580-022-00483-w.
14. To, T.K., and Kakutani, T. (2022). Crosstalk among pathways to generate DNA methylome. Curr. Opin. Plant Biol. *68*, 102248.
15. Murawska, M., Greenstein, R.A., Schauer, T., Olsen, K.C.F., Ng, H., Ladurner, A.G., Al-Sady, B., and Braun, S. (2021). The histone chaperone FACT facilitates heterochromatin spreading by regulating histone turnover and H3K9 methylation states. Cell Rep. *37*, 109944.
16. Jih, G., Iglesias, N., Currie, M.A., Bhanu, N.V., Paulo, J.A., Gygi, S.P., Garcia, B.A., and Moazed, D. (2017). Unique roles for histone H3K9me states in RNAi and heritable silencing of transcription. Nature *547*, 463–467.
17. Keenen, M.M., Brown, D., Brennan, L.D., Renger, R., Khoo, H., Carlson, C.R., Huang, B., Grill, S.W., Narlikar, G.J., and Redding, S. (2021). HP1 proteins compact DNA into mechanically and positionally stable phase separated domains. eLife *10*, e64563, https://doi.org/10.7554/eLife.64563.
18. Sanulli, S., Gross, J.D., and Narlikar, G.J. (2019). Biophysical properties of HP1-mediated heterochromatin. Cold Spring Harb. Symp. Quant. Biol. *84*, 217–225.
19. Chittock, E.C., Latwiel, S., Miller, T.C.R., and Müller, C.W. (2017). Molecular architecture of polycomb repressive complexes. Biochem. Soc. Trans. *45*, 193–205.
20. Holoch, D., and Margueron, R. (2017). Mechanisms regulating PRC2 recruitment and enzymatic activity. Trends Biochem. Sci. *42*, 531–542.
21. Verrijzer, C.P. (2022). Goldilocks meets Polycomb. Genes Dev. *36*, 1043–1045.
22. Fischer, S., Weber, L.M., and Liefke, R. (2022). Evolutionary adaptation of the Polycomb repressive complex 2. Epigenetics Chromatin *15*, 7.
23. Vijayanathan, M., Trejo-Arellano, M.G., and Mozgová, I. (2022). Polycomb repressive complex 2 in eukaryotes-an evolutionary perspective. Epigenomes *6*, 3, https://doi.org/10.3390/epigenomes6010003.
24. Blackledge, N.P., and Klose, R.J. (2021). The molecular principles of gene regulation by Polycomb repressive complexes. Nat. Rev. Mol. Cell Biol. *22*, 815–833.
25. Miró-Pina, C., Charmant, O., Kawaguchi, T., Holoch, D., Michaud, A., Cohen, I., Humbert, A., Jaszczyszyn, Y., Chevreux, G., Del Maestro, L., et al. (2022). Paramecium Polycomb repressive complex 2 physically interacts with the small RNA-binding PIWI protein to repress transposable elements. Dev. Cell *57*, 1037–1052.e8.
26. Frapporti, A., Miró Pina, C., Arnaiz, O., Holoch, D., Kawaguchi, T., Humbert, A., Eleftheriou, E., Lombard, B., Loew, D., Sperling, L., et al. (2019). The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in Paramecium. Nat. Commun. *10*, 2710.
27. Kramer, H.M., Seidl, M.F., Thomma, B.P.H.J., and Cook, D.E. (2021). Local rather than global H3K27me3 dynamics are associated with differential gene expression in Verticillium dahliae. mBio *13*, e0356621.
28. Carlier, F., Li, M., Maroc, L., Debuchy, R., Souaid, C., Noordermeer, D., Grognet, P., and Malagnac, F. (2021). Loss of EZH2-like or SU(VAR)3-9-like proteins causes simultaneous perturbations in H3K27 and H3K9 tri-methylation and associated developmental defects in the fungus Podospora anserina. Epigenetics Chromatin *14*, 22.
29. Mikulski, P., Komarynets, O., Fachinelli, F., Weber, A.P.M., and Schubert, D. (2017). Characterization of the polycomb-group mark H3K27me3 in unicellular algae. Front. Plant Sci. *8*, 607.
30. Schubert, D. (2019). Evolution of Polycomb-group function in the green lineage. F1000Res *8*, 8, https://doi.org/10.12688/f1000research.16986.1.
31. Montgomery, S.A., Tanizawa, Y., Galik, B., Wang, N., Ito, T., Mochizuki, T., Akimcheva, S., Bowman, J.L., Cognat, V., Maréchal-Drouard, L., et al. (2020). Chromatin organization in early land plants reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. Curr. Biol. *30*, 573–588.e7.
32. Zhao, X., Rastogi, A., Deton Cabanillas, A.F., Ait Mohamed, O., Cantrel, C., Lombard, B., Murik, O., Genovesio, A., Bowler, C., Bouyer, D., et al. (2021). Genome wide natural variation of H3K27me3 selectively marks

genes predicted to be important for cell differentiation in Phaeodactylum tricornutum. New Phytol. *229*, 3208–3220.

33. Déléris, A., Berger, F., and Duharcourt, S. (2021). Role of Polycomb in the control of transposable elements. Trends Genet. *37*, 882–889. https://doi.org/10.1016/j.tig.2021.06.003.
34. Yang, E.C., Boo, S.M., Bhattacharya, D., Saunders, G.W., Knoll, A.H., Fredericq, S., Graf, L., and Yoon, H.S. (2016). Divergence time estimates and the evolution of major lineages in the florideophyte red algae. Sci. Rep. *6*, 21361.
35. Huff, J.T., and Zilberman, D. (2014). Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell *156*, 1286–1297.
36. Zemach, A., Kim, M.Y., Hsieh, P.-H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., Harmer, S.L., and Zilberman, D. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. Cell *153*, 193–205.
37. Bourguet, P., Yelagandula, R., To, T.K., Osakabe, A., Alishe, A., Lu, R.J.-H., Kakutani, T., Chen, P.-Y., and Berger, F. (2022). The histone variant H2A.W cooperates with chromatin modifications and linker histone H1 to maintain transcriptional silencing of transposons in Arabidopsis. https://doi.org/10.1101/2022.05.31.493688.
38. Harris, B.J., Clark, J.W., Schrempf, D., Szöllősi, G.J., Donoghue, P.C.J., Hetherington, A.M., and Williams, T.A. (2022). Divergent evolutionary trajectories of bryophytes and tracheophytes from a complex common ancestor of land plants. Nat. Ecol. Evol. *6*, 1634–1643.
39. Widiez, T., Symeonidi, A., Luo, C., Lam, E., Lawton, M., and Rensing, S.A. (2014). The chromatin landscape of the moss Physcomitrella patens and its dynamics during development and drought stress. Plant J. *79*, 67–81.
40. Puttick, M.N., Morris, J.L., Williams, T.A., Cox, C.J., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Schneider, H., Pisani, D., et al. (2018). The interrelationships of land plants and the nature of the ancestral embryophyte. Curr. Biol. *28*, 733–745.e2.
41. Hisanaga, T., Wu, S., Axelsson, E., Akimcheva, S., Dolan, L., and Berger, F. (2022). The chromatin landscape of bryophytes. https://doi.org/10.1101/2022.10.21.513199.
42. Sugano, S.S., Nishihama, R., Shirakawa, M., Takagi, J., Matsuda, Y., Ishida, S., Shimada, T., Hara-Nishimura, I., Osakabe, K., and Kohchi, T. (2018). Efficient CRISPR/Cas9-based genome editing and its application to conditional genetic analysis in Marchantia polymorpha. PLoS One *13*, e0205117.
43. Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., et al. (2017). Insights into land plant evolution garnered from the Marchantia polymorpha genome. Cell *171*, 287–304.e15.
44. Montgomery, S.A., Hisanaga, T., Wang, N., Axelsson, E., Akimcheva, S., Sramek, M., Liu, C., and Berger, F. (2022). Polycomb-mediated repression of paternal chromosomes maintains haploid dosage in diploid embryos of Marchantia. eLife *11*, e79258, https://doi.org/10.7554/eLife.79258.
45. Flores-Sandoval, E., Dierschke, T., Fisher, T.J., and Bowman, J.L. (2016). Efficient and inducible use of artificial microRNAs in Marchantia polymorpha. Plant Cell Physiol. *57*, 281–290.
46. Spit, A., Hyland, R.H., Mellor, E.J., and Casselton, L.A. (1998). A role for heterodimerization in nuclear localization of a homeodomain protein. Proc. Natl. Acad. Sci. USA *95*, 6228–6233.
47. Lee, J.-H., Lin, H., Joo, S., and Goodenough, U. (2008). Early sexual origins of homeoprotein heterodimerization and evolution of the plant KNOX/BELL family. Cell *133*, 829–840.
48. Dierschke, T., Flores-Sandoval, E., Rast-Somssich, M.I., Althoff, F., Zachgo, S., and Bowman, J.L. (2021). Gamete expression of TALE class HD genes activates the diploid sporophyte program in Marchantia polymorpha. eLife *10*, e57088, https://doi.org/10.7554/eLife.57088.
49. Hisanaga, T., Fujimoto, S., Cui, Y., Sato, K., Sano, R., Yamaoka, S., Kohchi, T., Berger, F., and Nakajima, K. (2021). Deep evolutionary origin of gamete-directed zygote activation by KNOX/BELL transcription factors in green plants. eLife *10*, e57090, https://doi.org/10.7554/eLife.57090.
50. Ikeda, Y., Nishihama, R., Yamaoka, S., Arteaga-Vazquez, M.A., Aguilar-Cruz, A., Grimanelli, D., Pogorelcnik, R., Martienssen, R.A., Yamato, K.T., Kohchi, T., et al. (2018). Loss of CG methylation in Marchantia polymorpha causes disorganization of cell division and reveals unique DNA methylation regulatory mechanisms of non-CG methylation. Plant Cell Physiol. *59*, 2421–2431.
51. Jamge, B., Lorković, Z.J., Axelsson, E., Yelagandula, R., Akimcheva, S., and Berger, F. (2022). Transcriptional activity is shaped by the chromatin landscapes in Arabidopsis. https://doi.org/10.1101/2022.06.02.494419.
52. Wu, M.-F., Sang, Y., Bezhani, S., Yamaguchi, N., Han, S.-K., Li, Z., Su, Y., Slewinski, T.L., and Wagner, D. (2012). SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the LEAFY and SEPALLATA3 transcription factors. Proc. Natl. Acad. Sci. USA *109*, 3576–3581.
53. Sridhar, V.V., Surendrarao, A., and Liu, Z. (2006). APETALA1 and SEPALLATA3 interact with SEUSS to mediate transcription repression during flower development. Development *133*, 3159–3166.
54. Bezhani, S., Winter, C., Hershman, S., Wagner, J.D., Kennedy, J.F., Kwon, C.S., Pfluger, J., Su, Y., and Wagner, D. (2007). Unique, shared, and redundant roles for the Arabidopsis SWI/SNF chromatin remodeling ATPases BRAHMA and SPLAYED. Plant Cell *19*, 403–416.
55. Smaczniak, C., Immink, R.G.H., Muiño, J.M., Blanvillain, R., Busscher, M., Busscher-Lange, J., Dinh, Q.D.P., Liu, S., Westphal, A.H., Boeren, S., et al. (2012). Characterization of MADS-domain transcription factor complexes in Arabidopsis flower development. Proc. Natl. Acad. Sci. USA *109*, 1560–1565.
56. Pajoro, A., Madrigal, P., Muiño, J.M., Matus, J.T., Jin, J., Mecchia, M.A., Debernardi, J.M., Palatnik, J.F., Balazadeh, S., Arif, M., et al. (2014). Dynamics of chromatin accessibility and gene regulation by MADS-domain transcription factors in flower development. Genome Biol. *15*, R41.
57. Lai, X., Blanc-Mathieu, R., GrandVuillemin, L., Huang, Y., Stigliani, A., Lucas, J., Thévenon, E., Loue-Manifel, J., Turchi, L., Daher, H., et al. (2021). The LEAFY floral regulator displays pioneer transcription factor properties. Mol. Plant *14*, 829–837.
58. Jin, R., Klasfeld, S., Zhu, Y., Fernandez Garcia, M., Xiao, J., Han, S.-K., Konkol, A., and Wagner, D. (2021). LEAFY is a pioneer transcription factor and licenses cell reprogramming to floral fate. Nat. Commun. *12*, 626.
59. Zheng, S., Hu, H., Ren, H., Yang, Z., Qiu, Q., Qi, W., Liu, X., Chen, X., Cui, X., Li, S., et al. (2019). The Arabidopsis H3K27me3 demethylase JUMONJI 13 is a temperature and photoperiod dependent flowering repressor. Nat. Commun. *10*, 1303.
60. Huang, T.-C., Wang, Y.-F., Vazquez-Ferrer, E., Theofel, I., Requena, C.E., Hanna, C.W., Kelsey, G., and Hajkova, P. (2021). Sex-specific chromatin remodelling safeguards transcription in germ cells. Nature *600*, 737–742.
61. Rodriguez, F., Yushenova, I.A., DiCorpo, D., and Arkhipova, I.R. (2022). Bacterial N4-methylcytosine as an epigenetic mark in eukaryotic DNA. Nat. Commun. *13*, 1072.
62. Berke, L., and Snel, B. (2015). The plant Polycomb repressive complex 1 (PRC1) existed in the ancestor of seed plants and has a complex duplication history. BMC Evol. Biol. *15*, 44.
63. Sharaf, A., Vijayanathan, M., Oborník, M., and Mozgová, I. (2022). Phylogenetic profiling resolves early emergence of PRC2 and illuminates its functional core. Life Sci. Alliance *5*, 5. https://doi.org/10.26508/lsa.202101271.
64. Grau-Bové, X., Navarrete, C., Chiva, C., Pribasnig, T., Antó, M., Torruella, G., Galindo, L.J., Lang, B.F., Moreira, D., López-Garcia, P., et al. (2022). A phylogenetic and proteomic reconstruction of eukaryotic chromatin evolution. Nat. Ecol. Evol. *6*, 1007–1023.

65. Onishi, R., Yamanaka, S., and Siomi, M.C. (2021). piRNA- and siRNA-mediated transcriptional repression in Drosophila, mice, and yeast: new insights and biodiversity. EMBO Rep. *22*, e53062.

66. Zhou, Y., Wang, Y., Krause, K., Yang, T., Dongus, J.A., Zhang, Y., and Turck, F. (2018). Telobox motifs recruit CLF/SWN-PRC2 for H3K27me3 deposition via TRB factors in Arabidopsis. Nat. Genet. *50*, 638–644.

67. Chuong, E.B., Elde, N.C., and Feschotte, C. (2017). Regulatory activities of transposable elements: from conflicts to benefits. Nat. Rev. Genet. *18*, 71–86.

68. Batista, R.A., Moreno-Romero, J., Qiu, Y., van Boven, J., Santos-González, J., Figueiredo, D.D., and Köhler, C. (2019). The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. eLife *8*, e50541, https://doi.org/10.7554/eLife.50541.

69. Baud, A., Wan, M., Nouaud, D., Francillonne, N., Anxolabéhère, D., and Quesneville, H. (2022). Traces of transposable elements in genome dark matter co-opted by flowering gene regulation networks. Peer Community J. *2*, 2. https://doi.org/10.24072/pcjournal.68.

70. Roquis, D., Robertson, M., Yu, L., Thieme, M., Julkowska, M., and Bucher, E. (2021). Genomic impact of stress-induced transposable element mobility in Arabidopsis. Nucleic Acids Res. *49*, 10431–10447.

71. Thieme, M., Brêchet, A., Bourgeois, Y., Keller, B., Bucher, E., and Roulin, A.C. (2022). Experimentally heat-induced transposition increases drought tolerance in Arabidopsis thaliana. New Phytol. *236*, 182–194.

72. Baduel, P., Leduque, B., Ignace, A., Gy, I., Gil, J., Jr., Loudet, O., Colot, V., and Quadrana, L. (2021). Genetic and environmental modulation of transposition shapes the evolutionary potential of Arabidopsis thaliana. Genome Biol. *22*, 138.

73. Quadrana, L., Etcheverry, M., Gilly, A., Caillieux, E., Madoui, M.-A., Guy, J., Bortolini Silveira, A., Engelen, S., Baillet, V., Wincker, P., et al. (2019). Transposition favors the generation of large effect mutations that may facilitate rapid adaption. Nat. Commun. *10*, 3421.

74. Sammarco, I., Pieters, J., Salony, S., Toman, I., Zolotarov, G., and Lafon Placette, C. (2022). Epigenetic targeting of transposon relics: beating the dead horses of the genome? Epigenetics *17*, 1331–1344.

75. Baduel, P., and Quadrana, L. (2021). Jumpstarting evolution: how transposition can facilitate adaptation to rapid environmental changes. Curr. Opin. Plant Biol. *61*, 102043.

76. Studer, A., Zhao, Q., Ross-Ibarra, J., and Doebley, J. (2011). Identification of a functional transposon insertion in the maize domestication gene tb1. Nat. Genet. *43*, 1160–1163.

77. Butelli, E., Licciardello, C., Zhang, Y., Liu, J., Mackay, S., Bailey, P., Reforgiato-Recupero, G., and Martin, C. (2012). Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. Plant Cell *24*, 1242–1255.

78. Ong-Abdullah, M., Ordway, J.M., Jiang, N., Ooi, S.-E., Kok, S.-Y., Sarpan, N., Azimi, N., Hashim, A.T., Ishak, Z., Rosli, S.K., et al. (2015). Loss of karma transposon methylation underlies the mantled somaclonal variant of oil palm. Nature *525*, 533–537.

79. Zervudacki, J., Yu, A., Amesefe, D., Wang, J., Drouaud, J., Navarro, L., and Deleris, A. (2018). Transcriptional control and exploitation of an immune-responsive family of plant retrotransposons. EMBO J. *37*, e98482, https://doi.org/10.15252/embj.201798482.

80. Xiao, J., Jin, R., Yu, X., Shen, M., Wagner, J.D., Pai, A., Song, C., Zhuang, M., Klasfeld, S., He, C., et al. (2017). Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in Arabidopsis. Nat. Genet. *49*, 1546–1552.

81. Yuan, L., Song, X., Zhang, L., Yu, Y., Liang, Z., Lei, Y., Ruan, J., Tan, B., Liu, J., and Li, C. (2021). The transcriptional repressors VAL1 and VAL2 recruit PRC2 for genome-wide Polycomb silencing in Arabidopsis. Nucleic Acids Res. *49*, 98–113.

82. Wang, M., Zhong, Z., Gallego-Bartolomé, J., Feng, S., Shih, Y.-H., Liu, M., Zhou, J., Richey, J.C., Ng, C., Jami-Alahmadi, Y., et al. (2023). Arabidopsis TRB proteins function in H3K4me3 demethylation by recruiting JMJ14. Nat. Commun. *14*, 1736.

83. Friedrich, T., Faivre, L., Bäurle, I., and Schubert, D. (2019). Chromatin-based mechanisms of temperature memory in plants. Plant Cell Environ. *42*, 762–770.

84. Bieluszewski, T., Xiao, J., Yang, Y., and Wagner, D. (2021). PRC2 activity, recruitment, and silencing: a comparative perspective. Trends Plant Sci. *26*, 1186–1198.

85. Matsuzaki, M., Misumi, O., Shin, I.T., Maruyama, S., Takahara, M., Miyagishima, S.-Y., Mori, T., Nishida, K., Yagisawa, F., Nishida, K., et al. (2004). Genome sequence of the ultrasmall unicellular red alga Cyanidioschyzon merolae 10D. Nature *428*, 653–657.

86. Nozaki, H., Takano, H., Misumi, O., Terasawa, K., Matsuzaki, M., Maruyama, S., Nishida, K., Yagisawa, F., Yoshida, Y., Fujiwara, T., et al. (2007). A 100%-complete sequence reveals unusually simple genomic features in the hot-spring red alga Cyanidioschyzon merolae. BMC Biol. *5*, 28.

87. Li, F.-W., Nishiyama, T., Waller, M., Frangedakis, E., Keller, J., Li, Z., Fernandez-Pozo, N., Barker, M.S., Bennett, T., Blázquez, M.A., et al. (2020). Anthoceros genomes illuminate the origin of land plants and the unique biology of hornworts. Nat. Plants *6*, 259–272.

88. Song, L., Huang, S.-S.C., Wise, A., Castanon, R., Nery, J.R., Chen, H., Watanabe, M., Thomas, J., Bar-Joseph, Z., and Ecker, J.R. (2016). A transcription factor hierarchy defines an environmental stress response network. Science *354*, aag1550, https://doi.org/10.1126/science.aag1550.

89. Heyndrickx, K.S., Van de Velde, J., Wang, C., Weigel, D., and Vandepoele, K. (2014). A functional and evolutionary perspective on transcription factor binding in Arabidopsis thaliana. Plant Cell *26*, 3894–3910.

90. Ómaoiléidigh, D.S., Wuest, S.E., Rae, L., Raganelli, A., Ryan, P.T., Kwasniewska, K., Das, P., Lohan, A.J., Loftus, B., Graciet, E., et al. (2013). Control of reproductive floral organ identity specification in Arabidopsis by the C function regulator AGAMOUS. Plant Cell *25*, 2482–2503.

91. Nagel, D.H., Doherty, C.J., Pruneda-Paz, J.L., Schmitz, R.J., Ecker, J.R., and Kay, S.A. (2015). Genome-wide identification of CCA1 targets uncovers an expanded clock network in Arabidopsis. Proc. Natl. Acad. Sci. USA *112*, E4802–E4810.

92. Birkenbihl, R.P., Kracher, B., Roccaro, M., and Somssich, I.E. (2017). Induced genome-wide binding of three Arabidopsis WRKY transcription factors during early MAMP-triggered immunity. Plant Cell *29*, 20–38.

93. Obayashi, T., Aoki, Y., Tadaka, S., Kagaya, Y., and Kinoshita, K. (2018). ATTED-II in 2018: a plant coexpression database based on investigation of the statistical property of the mutual rank index. Plant Cell Physiol. *59*, e3.

94. Minoda, A., Sakagami, R., Yagisawa, F., Kuroiwa, T., and Tanaka, K. (2004). Improvement of culture conditions and evidence for nuclear transformation by homologous recombination in a red alga, Cyanidioschyzon merolae 10D. Plant Cell Physiol. *45*, 667–671.

95. Taki, K., Sone, T., Kobayashi, Y., Watanabe, S., Imamura, S., and Tanaka, K. (2015). Construction of a URA5.3 deletion strain of the unicellular red alga Cyanidioschyzon merolae: a backgroundless host strain for transformation experiments. J. Gen. Appl. Microbiol. *61*, 211–214.

96. Ishizaki, K., Nishihama, R., Yamato, K.T., and Kohchi, T. (2016). Molecular genetic tools and techniques for Marchantia polymorpha research. Plant Cell Physiol. *57*, 262–270.

97. Hisanaga, T., Okahashi, K., Yamaoka, S., Kajiwara, T., Nishihama, R., Shimamura, M., Yamato, K.T., Bowman, J.L., Kohchi, T., and Nakajima, K. (2019). A cis -acting bidirectional transcription switch controls sexual dimorphism in the liverwort. EMBO J. *38*, e100240, https://doi.org/10.15252/embj.2018100240.

98. Ou, S., Su, W., Liao, Y., Chougule, K., Agda, J.R.A., Hellinga, A.J., Lugo, C.S.B., Elliott, T.A., Ware, D., Peterson, T., et al. (2019). Benchmarking

transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. *20*, 275.

99. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. Bioinformatics *27*, 1571–1572.
100. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.
101. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.
102. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.
103. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.
104. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137.
105. Anand, L., and Rodriguez Lopez, C.M. (2022). ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. BMC Bioinformatics *23*, 33.
106. Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44*, W160–W165.
107. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
108. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
109. Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics *21*, 3448–3449.
110. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.
111. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer).
112. Yu, G., Wang, L.-G., and He, Q.-Y. (2015). ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. Bioinformatics *31*, 2382–2383.
113. Ewels, P., Hüther, P., Spix, N., Miller, E., Bot, N.-C., Hörtenhuber, M., Peltzer, A., Alneberg, J., Tommaso, P.D., Garcia, M.U., et al. (2022). nf-core/methylseq: nf-core/methylseq version 2.3.0 [pyrite Alligator]. https://doi.org/10.5281/zenodo.7449839.
114. Dorfer, V., Pichler, P., Stranzl, T., Stadlmann, J., Taus, T., Winkler, S., and Mechtler, K. (2014). MS Amanda, a universal identification algorithm optimized for high accuracy tandem mass spectra. J. Proteome Res. *13*, 3679–3684.
115. Käll, L., Canterbury, J.D., Weston, J., Noble, W.S., and MacCoss, M.J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat. Methods *4*, 923–925.
116. Taus, T., Köcher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., and Mechtler, K. (2011). Universal and confident phosphorylation site localization using phosphoRS. J. Proteome Res. *10*, 5354–5362.
117. Doblmann, J., Dusberger, F., Imre, R., Hudecz, O., Stanek, F., Mechtler, K., and Dürnberger, G. (2019). apQuant: accurate label-free quantification by quality filtering. J. Proteome Res. *18*, 535–541.
118. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.
119. Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.
120. Kobayashi, Y., Ohnuma, M., Kuroiwa, T., Tanaka, K., and Hanaoka, M. (2010). The basics of cultivation and molecular genetic analysis of the unicellular red alga Cyanidioschyzon merolae. Endocytobiosis Cell Res. *20*, 53–61.
121. Cho, C.H., Park, S.I., Huang, T.-Y., Lee, Y., Ciniglia, C., Yadavalli, H.C., Yang, S.W., Bhattacharya, D., and Yoon, H.S. (2023). Genome-wide signatures of adaptation to extreme environments in red algae. Nat. Commun. *14*, 10.
122. Imamura, S., Terashita, M., Ohnuma, M., Maruyama, S., Minoda, A., Weber, A.P.M., Inouye, T., Sekine, Y., Fujita, Y., Omata, T., et al. (2010). Nitrate assimilatory genes and their transcriptional regulation in a unicellular red alga Cyanidioschyzon merolae: genetic evidence for nitrite reduction by a sulfite reductase-like enzyme. Plant Cell Physiol. *51*, 707–717.
123. Tsuboyama, S., and Kodama, Y. (2018). AgarTrap protocols on your benchtop: simple methods for Agrobacterium-mediated genetic transformation of the liverwort Marchantia polymorpha. Plant Biotechnol. (Tokyo) *35*, 93–99.
124. Montgomery, S.A., and Berger, F. (2023). Epigenetic reprogramming of imprinting at meiosis. https://doi.org/10.1101/2023.05.17.541143.
125. Wang, Q., Gu, L., Adey, A., Radlwimmer, B., Wang, W., Hovestadt, V., Bähr, M., Wolf, S., Shendure, J., Eils, R., et al. (2013). Tagmentation-based whole-genome bisulfite sequencing. Nat. Protoc. *8*, 2022–2032.
126. Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinformatics *26*, 2204–2207.
127. Inoue, K., Nishihama, R., Kataoka, H., Hosaka, M., Manabe, R., Nomoto, M., Tada, Y., Ishizaki, K., and Kohchi, T. (2016). Phytochrome signaling is mediated by PHYTOCHROME INTERACTING FACTOR in the liverwort Marchantia polymorpha. Plant Cell *28*, 1406–1421.
128. Ernst, J., and Kellis, M. (2013). Interplay between chromatin state, regulator binding, and regulatory motifs in six human cell types. Genome Res. *23*, 1142–1154.
129. Ekwall, K., Olsson, T., Turner, B.M., Cranston, G., and Allshire, R.C. (1997). Transient inhibition of histone deacetylation alters the structural and functional imprint at fission yeast centromeres. Cell *91*, 1021–1032.
130. Blum, H., Beier, H., and Gross, H.J. (1987). Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels. Electrophoresis *8*, 93–99.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mouse monoclonal anti-H3	Diagenode	Cat# C15200011; Lot 002
Rabbit polyclonal anti-H3K9me1	Abcam	Cat# ab8896; Lot GR34167862; RRID: AB_732929
Rabbit polyclonal anti-H3K27me1	Millipore	Cat# 07-448; DAM1661077; RRID: AB_310623
Rabbit polyclonal anti-H3K27me3	Diagenode	Cat# C15410195; Lot A0821D; RRID: AB_2753161
IRDye 680 RD Goat anti-Mouse IgG Secondary Antibody	LI-COR	Cat# 926-68070; RRID: AB_10956588
IRDye 800 CW Goat anti-Rabbit IgG Secondary Antibody	LI-COR	Cat# 926-32211; RRID: AB_621843
Bacterial and virus strains		
Agrobacterium tumefaciens GV3101 (MP90)	Widely distributed	N/A
Escherichia coli DH5α	Widely distributed	N/A
Chemicals, peptides, and recombinant proteins		
Gamborg B5 basal salt	Duchefa	Cat# G0209
Hygromycin	Applichem	Cat# A2175
Chlorsulfuron	VWR	Cat# EHERC11610000
Cefotaxime sodium	Duchefa	Cat# C0111
3,5-Dimethoxy-4-hydroxyacetophenone (Acetosyringone)	Sigma-Aldrich	Cat#D134406
cOmplete, EDTA-free Protease Inhibitor Cocktail	Sigma-Aldrich	Cat# 11873580001
Critical commercial assays		
In-Fusion Cloning Kit	Takara Bio	Cat# 638910
INNUprep RNA Kit	Analytik Jena	Cat# 845-KS-20800050
DNase I	Thermo Fisher Scientific	Cat# EN0521
RiboLock RNase inhibitor	Thermo Fisher Scientific	Cat# EO0381
NEBNext Ultra RNA Library Prep Kit for Illumina	New England Biolabs	Cat# E7530L
BsaI HFv2	New England Biolabs	Cat# R3733L
T4 DNA Ligase	New England Biolabs	Cat# M0202L
BglI	Thermo Fisher Scientific	Cat# ER0071
Gateway LR Clonase II Enzyme mix	Thermo Fisher Scientific	Cat#11791-020
Spectrum Plant Total RNA kit	Sigma-Aldrich	STRN250
DNA-free DNA Removal Kit	Thermo Fisher Scientific	AM1906
NEBNext Poly(A) mRNA Magnetic Isolation Module	New England Biolabs	E7490L
NEBNext Ultra II Directional RNA Library Prep Kit for Illumina	New England Biolabs	E7760L
RevertAid H Minus First Strand cDNA Synthesis Kit	Thermo Fisher Scientific	K1632
Luna Universal qPCR Master Mix	New England Biolabs	M3003E
Immobilon-FL PVDF	Merck Millipore	IPFL00010

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Cyanidioschyzon merolae genome ASM9120v1	Matsuzaki et al.[85]; Nozaki et al.[86]	http://czon.jp/
Bisulfite-seq data of *Cyanidioschyzon merolae*	Huff et al.[35]	GEO: GSE46692
ChIP-seq data of *Cyanidioschyzon merolae*	Mikulski et al.[29]	GEO: GSE93913
RNA-seq data of Cme*(z)-1* and WT for expression analysis	This paper	GEO: GSE221632
Anthoceros agrestis Oxford strain genome	Hisanaga et al.[41]; Li et al.[87]	GEO: GSE218880
ChIP-seq data of *Anthoceros agrestis*	Hisanaga et al.[41]	GEO: GSE218878
Marchantia polymorpha genome v5.1	Montgomery et al.[31]	http://marchantia.info
CUT&RUN data of *Marchantia polymorpha*	Montgomery et al.[31]	N/A
RNA-seq data of Mpe*(z)1* mutants for expression analysis	This paper	GEO: GSE221631
ChIP-seq data of *Arabidopsis thaliana* chromatin modifications	Jamge et al.[51]	N/A
ATAC-seq data of *Arabidopsus thaliana*	Jamge et al.[51]	N/A
ChIP-seq data of *Arabidopsis thaliana* transcription factors	Song et al.[88]; Heyndrickx et al.[89]; ÓMaoiléidigh et al.[90]; Nagel et al.[91]; Birkenbihl et al.[92]	N/A
RNA-seq data of Mp*met* mutant	This paper	GEO: GSE234509
Bisulfite-seq data of Mpmet	This paper	GEO: GSE234851
Transcriptome data of *Phaeodactylum tricornutum* Pte*(z)* mutant	Zhao et al.[32]	SRA: PRJNA565539
coexpressed gene tables	Obayashi et al.[93]; ATTEDII version Ath-r.c3-1	https://atted.jp/
RNA-seq data of *Arabidopsis thaliana ddm1* and *met1* mutants	Zemach et al.[36]	SRA: SRR578941, SRR578942, SRR578945-SRR578948
RNA-seq data of *Arabidopsis thaliana suvh4/5/6 mutant*	Bourguet et al.[37]	N/A
Experimental models: Organisms/strains		
Cyanidioschyzon merolae 10D	Matsuzaki et al.[85]; Nozaki et al.[86]	N/A
Cyanidioschyzon merolae M4	Minoda et al.[94]	N/A
Cyanidioschyzon merolae Cme*(z)-1*	This paper	N/A
Cyanidioschyzon merolae T1	Taki et al.[95]	N/A
Cyanidioschyzon merolae Cme*(z)-2*	This paper	N/A
Marchantia polymorpha Tak-1	Ishizaki et al.[96]	N/A
Marchantia polymorpha Tak-2	Ishizaki et al.[96]	N/A
Marchantia polymorpha Mp*knox2-1*	This paper	N/A
Marchantia polymorpha Mpknox2-1 Mpe*(z)1-1*	This paper	N/A
Marchantia polymorpha Mpknox2-1 Mpe*(z)1-2*	This paper	N/A
Marchantia polymorpha Mpknox2-2 Mpe*(z)1-3*	This paper	N/A
Marchantia polymorpha Mp*met-3*	Ikeda et al.[50]	N/A
Oligonucleotides		
See Table S3	N/A	N/A
Recombinant DNA		
pUC19	Widely distributed	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pSR875	Kind gift from Stephen Rader and Martha Stark (University of Northern British Columbia, Canada)	N/A
pMpGE_En03	Addgene	Cat# 71535; RRID: Addgene_71535
pMpGE010	Addgene	Cat# 71536; RRID: Addgene_71536
pMpGE_En04	Hisanaga et al.[97]	N/A
pBC-GE14	Hisanaga et al.[97]	N/A
pMpGE011	Addgene	Cat# 71537; RRID: Addgene_71537
pMpGE_En03-MpKNOX2ge1	This paper	N/A
pMpGE010_MpKNOX2ge1	This paper	N/A
pMpGE_En04-MpEz1ge1	This paper	N/A
pBC-GE14-MpEz1ge4	This paper	N/A
pMpGE_En04-MpEz1-ge1-ge4	This paper	N/A
pMpGE011_MpEz1-ge1-ge4	This paper	N/A
pBC-GE14-MpKNOX2ge1	This paper	N/A
pMpGE_En04-MpEz1-ge1-MpKNOX2ge1	This paper	N/A
pMpGE010_MpEz1ge1-MpKNOX2ge1	This paper	N/A
Software and algorithms		
EDTA 1.9.9	Ou et al.[98]	https://github.com/oushujun/EDTA
Trim Galore	Babraham Institute	https://github.com/FelixKrueger/TrimGalore
Bismark v0.22.2	Krueger and Andrews[99]	https://www.bioinformatics.babraham.ac.uk/projects/bismark/
Bowtie2 v2.3.4.2	Langmead and Salzberg[100]	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
SAMtools v1.9	Li et al.[101]	http://www.htslib.org/
BEDTools v2.27.1	Quinlan and Hall[102]	https://bedtools.readthedocs.io/en/latest/
Cutadapt v1.18	Martin[103]	N/A
Picard v2.18.27	Broad Institute, Boston, MA	http://broadinstitute.github.io/picard/
macs2 v2.2.5	Zhang et al.[104]	https://github.com/macs3-project/MACS
R v1.3, v4.2.0	R Foundation for Statistical Computing, Vienna, Austria	https://www.R-project.org/
Chromomap	Anand and Rodriguez Lopez[105]	https://lakshay-anand.github.io/chromoMap/docs.html
deepTools v3.3.1	Ramírez et al.[106]	https://deeptools.readthedocs.io/en/develop/
STAR	Dobin et al.[107]	https://github.com/alexdobin/STAR
TELocal	Hammell lab, Cold Spring Harbor Laboratory	https://github.com/mhammell-laboratory/TElocal
DeSeq2 v1.22.2	Love et al.[108]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
BiNGO	Maere et al.[109]	https://www.psb.ugent.be/cbd/papers/BiNGO/Home.html
Cytoscape	Shannon et al.[110]	https://cytoscape.org/
ggplot2	Wickham[111]	https://ggplot2.tidyverse.org/
ChIPseeker	Yu et al.[112]	https://bioconductor.org/packages/release/bioc/html/ChIPseeker.html
nf-core/methylseq v2.3.0	Ewels et al.[113]	https://nf-co.re/methylseq/2.3.0
Proteome Discoverer version 2.5.0.400	Thermo Scientific	N/A
MSAmanda v2.0.0.19924	Dorfer et al.[114]	https://ms.imp.ac.at/?goto=msamanda
Percolator algorithm	Käll et al.[115]	N/A
phosphoRS	Taus et al.[116]	N/A

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
IMP-apQuant	Doblmann et al.[117]	N/A
iBAQ	Schwanhäusser et al.[118]	N/A
Fiji	Schindelin et al.[119]	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Frédéric Berger (frederic.berger@gmi.oeaw.ac.at).

Materials availability

Plasmids, *C. merolae* lines and *M. polymorpha* lines generated in this study are available upon request.

Data and code availability

High through-put sequencing data generated in this study have been deposited at NCBI Gene Expression Omnibus and are publicly available as of the date of publication. Accession numbers are listed in the key resources table. This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. All code used in this study is available upon request. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Plant materials

Marchantia polymorpha L. subsp. *ruderalis* accessions Takaragaike 1 (Tak-1) and Takaragaike 2 (Tak-2)[96] were used as the wild-type male and female, respectively. Mpmet-3 mutant was a kind gift from Dr. Yoko Ikeda (Okayama University, Japan) and was described in Ikeda et al.[50] Plants were cultured on half-strength Gamborg's B5 medium solidified with 1% (w/v) agar under continuous white light at 22°C.

Alga materials

Cyanidioschyzon merolae 10D and Cme*(z)* cells were grown in liquid culture in sterile modified 2x concentrated Allen's medium (MA2,[120]) at 2.5 < pH < 3.0 under constant white light (80 μmol/m^2/s) at 42°C. Cultures were kept in 50 ml falcon tubes aerated with ambient air supplied through a 1 ml serological milk pipette coupled to an aquarium pump. No additional CO_2 was supplied.

METHOD DETAILS

Annotation of TEs in *C. merolae*, *A. agrestis* and *M. polymorpha*

Transposable elements (TEs) of *C. merolae* and *M. polymorpha* were annotated using EDTA 1.9.9,[98] which incorporates a bunch of tools including LTRharvest, LTR_FINDER, LTR_retriever, Generic Repeat Finder, TIR-Learner, MITE-Hunter, HelitronScanner, and RepeatMasker. All softwares are adjusted to EDTA with proper filters and parameters. Final non-redundant TE libraries are produced by removing nested insertions and protein-coding genes by EDTA customized scripts. For *C. merolae*, we used a custom repeat library optimized for red algal TEs from a previous study[121] to classify predicted TEs using EDTA (Data S1). The TE classification for *M. polymorpha* was annotated based on the default parameters of EDTA (Data S1).

Re-analyses of DNA methylation in *C. merolae*

Bisulfite-seq data of *C. merolae* were downloaded from the sequence read archive of NCBI under the study PRJNA201680.[35] Reads were trimmed with Trim Galore (https://github.com/FelixKrueger/TrimGalore). Bisulfite converted reference genome was prepared from *C. merolae* ASM9120v1 genome sequence using Bismark v0.22.2.[99] Trimmed reads were mapped to the bisulfite genome using Bowtie2 v2.3.4.2[100] option of Bismark. Duplicates were removed using deduplicate function in Bismark. Cytosine methylation reports were created from deduplicated reads using bismark_methylation_extractor function in Bismark.

Generation of the Cme*(z)* mutant

To inactivate the Cm*E(z)* (CMQ156C) gene, the chromosomal Cm*E(z)* open reading frame (ORF) was replaced by the *C. merolae URA5.3* selectable marker gene by homologous recombination as follows. All the primers used are listed in Table S3.

To obtain Cm*e(z)-1*, the Cm*E(z)* genomic region, which contained the Cm*E(z)* ORF and its 1-kb each of 5′- and 3′-flanking sequences, was amplified with the primers E(z)_KO_F1 and E(z)_KO_R1. The amplified DNA was cloned into the vector pUC19 by In-Fusion Cloning Kit (Takara Bio, Japan). The 5'-flanking sequence of Cm*E(z)* ORF, the vector, and the 3'-flanking sequence of Cm*E(z)* ORF were amplified with the primers E(z)-KO_F2 and E(z)_KO_R2 and then the *URA5.3* gene, which was amplified with the primers URA_F and URA_R was inserted between the 5'- and 3'-flanking sequence of Cm*E(z)* ORF by In-Fusion Cloning Kit. The Cm*E(z)* genomic region, in which Cm*E(z)* ORF was replaced with *URA5.3*, was amplified with the primers pUC19_F and pUC19_R and was transformed into *C. merolae* M4, a derivative of *C. merolae* 10D, which has a mutation in the *URA5.3* gene.[94]

To obtain Cm*e(z)*-2 mutant, 0.5 kb of each, 5′and 3′-sequences, flanking the Cm*E(z)* ORF were amplified using the primer sets 5′UTR_E(z)_for/ 5′UTR_E(z)_rev and 3′UTR_E(z)_for/ 3′UTR_E(z)_rev, respectively. The amplified DNA of the 5′ and 3′ flanking region (putative untranslated region, UTR) was cloned into the *SwaI* and *PacI* site, respectively, of the plasmid pSR875 via ligation independent cloning (LIC) method. pSR875 was a kind gift of Stephen Rader and Martha Stark (University of Northern British Columbia, Canada). pSR875 is made from the pBS backbone with *SwaI* and *PacI* LIC sites added + a 10xHis tag + Nos terminator + the Ura5.3 cassette of *C. merolae*. The Cm*E(z)* genomic region, in which Cm*E(z)* ORF was replaced with *URA5.3*, was amplified with the primers Cm_trafo_A and Cm_trafo_B was transformed into *C. merolae* T1, a derivative of *C. merolae* 10D, which has a deleted *URA5.3* gene.[95] Transformation and selection of the gene knockouts were performed as described.[122]

Generation of transcriptome of *C. merolae*

C. merolae cells were sampled as follows: 2 ml of culture grown to OD_{750nm} = 1 were harvested by centrifugation 3000 x g for 3min at 4°C. The supernatant was removed and the pellet-containing reaction tube put into liquid nitrogen for 15 seconds for cell homogenization purposes. RNA was isolated from frozen pellets using the INNUprep RNA Kit (Analytik Jena) according to the manufacturerś instructions with the RL buffer supplemented with 10μl/ml beta-mercaptoethanol (β-ME) to improve RNA quality. RNA was eluted in RNAse-free water. Genomic DNA was removed from 1 μg prepared RNA via treatment with DNAse I (Thermo Fisher Scientific) according to manufacturerś instructions. RiboLock RNase inhibitor (Thermo Fisher Scientific) was added to a final concentration of 1 u/μl. mRNA library was prepared using polyA enrichment and the NEBnext Ultra RNA Library Prep Kit producing unstranded data. Sequencing was done on an Illumina NovaSeq 6000 platform resulting on 150 bp long paired-end reads. 3G of raw data per sample was produced.

ChIP-seq data analyses

Details for preparation for ChIP-seq libraries of *Anthoceros agrestis* are in.[41] ChIP-seq data of *C. merolae* were downloaded from the Gene expression omnibus of NCBI under the series GSE93913.[29] The bam files of ChIP-seq reads were sorted with SAMtools v1.9[101] and converted to fastq format using bamtofastq function of BEDTools v2.27.1,[102] and then trimmed with Cutadapt v1.18[103] and aligned to *A. agrestis* Oxford strain genome[41,87] or *C. merolae 10D* genome (ASM9120v1)[85,86] using Bowtie2 v2.3.4.2.[100] Resulting bam files were sorted and indexed with SAMtools v1.9. Reads with MAPQ less than ten were removed with Samtools v1.9 and duplicates were removed with Picard v2.18.27 (http://broadinstitute.github.io/picard/). Deduplicated reads from 2 (for *A. agrestis*) or 3 (for *C. merolae*) biological replicates were merged.

H3K27me3 broad peaks of *C. merolae* were called by using macs2 v2.2.5.[104] Coverage of H3K27me3 over PCGs and TEs are calculated using the intersect function of BEDtools v2.27.1. PCGs and TEs are considered as covered by H3K27me3 when more than 50% of the regions of each PCG or TE are overlapped by H3K27me3 peaks. The read coverage of H3K27me3 mark in *C. merolae* was normalized against the read coverage of H3 with bamCompare function in deepTools v3.3.1,[106] generating bigwig files.

The read coverage of each chromatin mark in *A. agrestis* was normalized against the read coverage of H3 with bamCompare function in deepTools v3.3.1,[106] generating bigwig files. K-means clustering of chromatin marks was performed using deepTools v3.3.1. Matrices were computed using computeMatrix for either PCGs or TEs using bigwig files as input. The reference-point sub-command was used to calculate scores of each mark over 2 kb upstream and downstream regions around the start codon of each PCG. The scale-regions sub-command was used to calculate scores of each mark over each TE scaled to 1 kb with 1 kb upstream and downstream. Aggregate profile plots of matrices were plotted with plotProfile with k-means clustering. Cluster assignments can be found in Data S4. Closest function in BEDTools v2.27.1 was used to define the closest TE and PCG pair.

Clustering analysis of chromatin marks of *M. polymorpha*

Bigwig files of five chromatin marks (H3K4me3, H3K9me1, H3K27me1, H3K27me3 and H3K36me3) in *M. polymorpha* were obtained from CUT&RUN datasets compiled previously.[31] K-means clustering of these five chromatin marks was performed using deepTools v3.3.1. Matrices were computed using computeMatrix for either PCGs (MpTak1v5.1_r2) or TEs annotated in this study using bigwig files as input. The reference-point sub-command was used to calculate scores of each mark over 2 kb upstream and downstream regions around the transcription start site (TSS) of each PCG. The scale-regions sub-command was used to calculate scores of each mark over each TE scaled to 1 kb with 1 kb upstream and downstream. Aggregate profile plots of matrices were plotted with plotProfile with k-means clustering. Cluster assignments can be found in Data S4. Closest function in BEDTools v2.27.1 was used to define the closest TE and PCG pair.

Generation of *M. polymorpha* PRC2 knockout mutant

All the primers used to generate *M. polymorpha* PRC2 knockout mutants are listed in Table S3. A DNA fragment producing MpKNOX2-targeting gRNAs was prepared by annealing a pair of synthetic oligonucleotides (TH637/TH638). The fragment was inserted into the BsaI site of pMpGE_En03 (cat. no. 71535, Addgene, Cambridge, MA) to yield pMpGE_En03-MpKNOX2ge1, which was transferred into pMpGE010 (cat. no. 71536, Addgene)[42] using the Gateway LR reaction (Thermo Fisher Scientific, Waltham, MA) to generate pMpGE010_MpKNOX2ge1. This construct was introduced into Tak-2 gemmae using the G-AgarTrap method.[123] Transformants were selected on 0.5 Gamborg B5 plates without vitamins (Duchefa Biochemie) supplemented with hygromycin and genotyped using the primer pair TH652/TH653, leading to isolation of Mp*knox2-1* allele.

To construct a plasmid to disrupt Mp*E(z)1*, two DNA fragments producing Mp*E(z)1*-targeting gRNAs were prepared by annealing pairs of synthetic oligonucleotides (MpEz1-gRNA-1-Fw/MpEz1-gRNA-1-Rv and MpEz1-gRNA-4-Fw/MpEz1-gRNA-4-Rv). The fragments were inserted into the BsaI sites of pMpGE_En04 and pBC-GE14[97] to yield pMpGE_En04-MpEz1ge1 and pBC-GE14-MpEz1ge4, respectively. These two plasmids were assembled via BglI restriction sites and ligated to yield pMpGE_En04-MpEz1-ge1-ge4. The resulting DNA fragment containing two MpU6promoter-gRNA cassettes was transferred into pMpGE011 (cat. no. 71537, Addgene) using the Gateway LR reaction (Thermo Fisher Scientific) to yield pMpGE011_MpEz1-ge1-ge4. This construct was introduced into Mp*knox2-1* gemmae using the G-AgarTrap method. Transformants were selected for on 0.5 Gamborg B5 plates without vitamins (Duchefa Biochemie) supplemented with chlorsulfuron and genotyped using the primer pair TH650/TH651, leading to isolation of Mp*knox2-1* Mp*e(z)1-1* and Mp*knox2-1*Mp*e(z)1-2* alleles.

To construct a plasmid to disrupt Mp*KNOX2* and Mp*E(z)1* simultaneously, a DNA fragment producing Mp*KNOX2*-targeting gRNAs was prepared by annealing a pair of synthetic oligonucleotides (TH637/TH638). The fragment was inserted into the BsaI site of pBC-GE14 to yield pBC-GE14-MpKNOX2ge1. This plasmid was assembled with pMpGE_En04-MpEz1ge1 via BglI restriction sites and ligated to yield pMpGE_En04-MpEz1-ge1-MpKNOX2ge1. The resulting DNA fragment containing two MpU6promoter-gRNA cassettes was transferred into pMpGE010 (cat. no. 71536, Addgene) using the Gateway LR reaction (Thermo Fisher Scientific) to yield pMpGE010_MpEz1ge1-MpKNOX2ge1. This construct was introduced into Tak-1 gemmae using the G-AgarTrap method. Transformants were selected for on 0.5 Gamborg B5 plates without vitamins (Duchefa Biochemie) supplemented with hygromycin and genotyped using the following primer pairs: TH650/TH651 for Mp*E(z)1*, TH652/TH653 for Mp*KNOX2*, leading to isolation of Mp*knox2-2* Mp*e(z)1-3*. To quantify the growth phenotype of Mp*e(z)* mutants, gemmae of the male wild type, Mp*knox2-2* Mp*e(z)1-3*, female wild type, Mp*knox2-1* and Mp*knox2-1* Mp*e(z)1-2* were grown on 1/2 Gamborg's B5 plates for 14 days, then photos of plants were taken by using Canon EOS 80D digital camera. Area of thalli was measured by using Fiji[119] and was plotted using R v4.2.0.

Generation of transcriptome of *M. polymorpha* PRC2 mutants and the Mpmet mutant

For the wild type, Mp*knox2* and Mp*knox2* Mp*e(z)1*, 14 day old plants grown from gemmae were collected. For Mp*met*, 14 day old plants grown from freshly transferred thallus pieces were collected. Samples were frozen in liquid nitrogen in Precellys tubes (Bertin Instruments, Montigny-le-Bretonneux, France) with 2.8mm zirconium oxide beads (Bertin, Rockville, MD, USA) and disrupted with a Precellys Evolution tissue homogenizer (Bertin Technologies) using the following settings: 4500 RPM 30s, 5s pause, repeated twice. Total RNA was extracted using a Spectrum Plant Total RNA kit (Sigma Aldrich, Merck KGaA, Darmstadt, Germany). Extracted RNA was treated by DNA-free DNA Removal Kit (Thermo Fisher Scientific). RNA-seq libraries were generated from 1 μg of total RNA using NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs) and NEBNext Ultra II Directional RNA Library Prep Kit for Illumina (New England Biolabs). These libraries were sequenced on an Illumina NextSeq 550 to generate 75bp paired-end reads or on a Illumina NextSeq 2000 to generate 50bp paired-end reads. Three biological replicates each of Mp*knox2-1* and Mp*knox2-1* Mp*e(z)1-1*, three biological replicates each of Mp*knox2-1* and Mp*knox2-1* Mp*e(z)1-2* and three biological replicates each of Tak-1 and Mpmet-3 were used for subsequent analyses.

Transcriptome data analysis

Bam files of RNA-seq reads were sorted with SAMtools v1.9[101] and converted to fastq format using bamtofastq function of BEDTools v2.27.1,[102] and then trimmed with Trim Galore (https://github.com/FelixKrueger/TrimGalore) and aligned to MpTak1v5.1r2 genome[31] for *M. polymorpha* or *C. merolae* ASM9120v1 genome for *C. merolae* using STAR.[107] Reads counts for PCGs and TEs were calculated by using TElocal (https://github.com/mhammell-laboratory/TElocal). Calculated read counts were imported into R v4.2.0 and differential gene expression analysis was performed using DESeq2 v1.22.2.[108] Estimated read counts of each PCG or TE and results of differential expression analyses in *C. merolae* and *M. polymorpha* can be found in Data S3, respectively. Estimated read counts and results of differential expression analyses of *P. tricornutum* were compiled previously.[32] Bigwig files of *C. merolae* RNA-seq data were generated from aligned bam files using bamCoverage function of deepTools v3.3.1 with normalization option –scaleFactor. The inverse of size factors calculated by DEseq2 was used as a scale factor for each replicate. Chromosomal plots in Figures 1F and 2H were generated using chromoMap v4.1.1[105] in R v4.2.0. PCGs and TEs commonly deregulated between Mp*knox2-1* Mp*e(z)1-1* and Mp*knox2-1* Mp*e(z)1-2* were used for downstream analyses. GO term enrichment analysis was done using BiNGO v3.0.5[109] plugin in Cytoscape v3.9.0.[110] To plot Figure 4F, we first selected pairs of TEs and their nearest PCGs covered by H3K27me3 by filtering the output of the BEDtools closest function described above with clusters Mp T1 and Mp P1. Then we identified expression states of PCGs and their nearest TEs in the Mp*knox2-1* Mp*e(z)1-1* mutant compared to the wild type using DESeq2 v1.22.2 as up if $\log_2$ fold change > 1 and adjusted P value < 0.05, down if $\log_2$ fold change < -1 and adjusted P value < 0.05 or unchanged for others.

This resulted in a total of 18 combinations (3 states of PCGs x 3 states of TEs x 2 (PCG/TE)). The numbers of TEs and PCGs belonging to each state were plotted into a heatmap using the ggplot2[111] package in R v4.2.0.

Transcriptome data of the *ddm1* and *met1* mutants in *A. thaliana* were obtained from a previous study[36] (SRR578941, SRR578942, SRR578945-SRR578948). Reads were trimmed, mapped to the *A. thaliana* Tair10 genome and then counted using the same method as described above. Read counts from transcriptome analyses of the *suvh4/5/6* mutant were obtained from Bourguet et al.[37] These counts data were used in differential gene expression analysis by DESeq2 v1.22.2. TEs with $\log_2$ fold change > 1 and adjusted P value < 0.05 were assigned as increased TEs. Proportions of increased TEs per TE cluster were calculated using TE cluster assignment in the previous study.[51]

Real time RT-PCR

Total RNAs were prepared from 14 day old plants grown from gemmae of Tak-2, Mp*knox2-1* and Mp*knox2-1* Mp*e(z)1-2* following the protocol described above. Extracted RNA was treated by DNA-free DNA Removal Kit (Thermo Fisher Scientific). cDNA was synthesized from 0.5 μg of total RNA using RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Fisher Scientific). Real-time RT-PCR analysis was performed using LightCycler 96 (Roche) and Luna Universal qPCR Master Mix (New England Biolabs) with the primers listed in Table S3. The analysis was done with three technical replicates and three biological replicates for each genotype. Expression levels of each TEs were normalized against the expression level of MpEF1α. Expression levels of each TE in each genotype were then normalized with the expression level in Mp*knox2-1* Mp*e(z)1-2* and plotted using the ggplot2 package in R v4.2.0.

Analyses of DNA methylation in M.polymorpha

For the wild type, 14 day old plants grown from gemmae were collected. For Mp*met*, 14 day old plants grown from freshly transferred thallus pieces were collected. These samples were used for whole-genome bisulfite sequencing by tagmentation described in Montgomery and Berger and Wang et al.[124,125] Prepared libraries were sequenced on a Illumina NextSeq 2000 to generate 50bp paired-end reads.

The fastq files of bisulfite-seq reads were trimmed, mapped and deduplicated to generate cytosine methylation reports using nf-core/methylseq v2.3.0 pipeline.[113] Each cytosine which is covered by at least six reads was used for further analyses. The methylation ratio of each cytosine was calculated and summarized to a bed file. These bed files were converted to bigwig files using bedGraphToBigWig[126] and used as inputs for the computMatrix function in deepTools v3.3.1. Aggregate profile plots of matrices were plotted with the plotProfile function in deepTools v3.3.1.

Nuclei isolation from M. polymorpha

To isolate nuclei from vegetative tissue of Marchantia, a method described in a previous study was used with some modifications.[127] 500 mg of thallus tissue from 14 day old plants grown from gemmae was collected in a 15 ml plastic tube and frozen with 6 mm zirconium beads. Frozen tissue was disrupted by vortex (max speed, 30 second repeat 6 times). Disrupted tissue was mixed with 5 ml of lysis buffer (20 mM Tris-HCl pH8.0, 25%[v/v] glycerol, 10 mM $MgCl_2$, 250mM sucrose, 5 mM DTT, 1x cOmplete) using vortex and then filtered through double-layered Miracloth. The flow-through was centrifuged at 1500g for 10 min at 4°C. The supernatant was discarded and the pellet was washed three times in 5 mL of wash buffer (20 mM Tris-HCl pH8.0, 25%[v/v] glycerol, 10 mM $MgCl_2$, 0.2% [v/v] TritonX-100, 5 mM DTT, 1x cOmplete). The final pellet was resuspended in 200 μl 1x Laemmli buffer in 0.2x PBS and boiled at 95°C for 5min.

Protein extraction from C. merolae and immunoblot analyses

Algae were grown to OD750=1 and concentrated/diluted accordingly. After centrifugation for 5 min at 10000 g, the pellet was dissolved in 100 μl of 4 M Urea, 100 μl of 2xSDS Loading buffer was added and proteins were denatured at 95 °C for 10 min. After separation of proteins on 15 % SDS-gels, they were blotted using Immobilon-FL PVDF membranes (IPFL00010, Merck Millipore). After transfer and activation the membrane was blocked with LiCor Odyssey Blocking Solution. Histone modification specific antibodies (α-H3K27me3, Diagenode C15410195, Lot A0821D, 1:2000; α-H3K9me1, Abcam 8896, Lot GR34167862, 3 μg; α-H3K27me1, Merck Millipore 07-448, Lot DAM1661077, 2 μg) were incubated in PBS Odyssey Blocking solution at 4°C overnight with the membrane, then anti-H3pan (Diagenode, C15200011, Lot 002, 1:2000) antibodies were added for 1h at room temperature. Membranes were washed and incubated with the two secondary antibodies (IRDye 680 RD, LiCor926-68070, 1:150000; IRDye 800 CW, Li-Cor 926-32211, 1:75000) for 1 h at room temperature. After drying of membrane, fluorescent signals were detected in a LiCor Odyssey XF at 700 nm and 800 nm, and quantified with LiCor Empiria Studio Software. For peptide competition, 10 μg peptides (H3K9me1, Diagenode C16000065; H3K27me1, Diagenode C16000045) were incubated with histone modification specific antibody in LiCor Odyssey blocking solution for 30 min at room temperature. Subsequently, the membrane was added and incubated overnight at 4°C.

TE clusters in A. thaliana

Transposable element (TE) cluster BED files as previously described in *A. thaliana* genome were obtained from Jamge et al.[51] For each TE cluster, comparisons were made in terms of their length, their genomic proportion, the TE family, and the heterochromatic marks associated with these TE clusters. Bigwig files for various heterochromatic marks (ChIP-seq) and ATAC-seq were downloaded from data compiled previously.[51] For each TE, the distance to the closest gene was calculated in R. The box plot and density

distribution plots were generated in R. The aggregate profile plots for ChIP-seq and ATAC-seq were generated using deepTools v3.3.1.[106]

TEG clusters in A. thaliana

To identify Transposable element genes (TEG) that were associated with each chromatin state as described in Jamge et al.,[51] we first converted the state region BED file into a signal bigwig file. Presence of state was assigned score "1" and absence of state was assigned "0". This scored signal file for each state was then used in Deeptools v3.3.1 to generate clusters of TEGs annotated in the *A. thaliana* TAIR10 genome based on state enrichment. Not all the states overlapped with TEGs, Hence only heterochromatic states (H1-H6), I1 and F2 were used for clustering of TEGs. We chose to analyze 3 unique groups of TEGs. The largest group TEG1 (n=2048) was covered with heterochromatic state H3 (enriched in all repressive histone marks such as H3K9me1/2, H2A.W.6/H2A.W.7 and H3K4me1), followed by TEG2 (n=991) covered by state H2 (enriched in similar repressive marks as in TEG1 with exception of H3K4me1), and third group TEG3 (n=864) was a mixed group of TEGs covered with multiple states in H4 and H6 and F2(enriched for repressive marks H3K9me1, H2A.W.7 or H3K27me3). A text file describing assignment of the TE and TEG clusters is provided in Data S5.

TF analysis in A. thaliana

Transcription factor peak BED files were obtained from ChIP-seq datasets compiled previously.[88–92] Each significant peak region in the genome was classified according to TE and TEG clusters.[51] Only the center of the peak was considered to classify chromatin states. Enrichment for TF binding was performed as in Ernst and Kellis[128] but replacing chromatin states for TEs from each cluster. Briefly, for each TE cluster (te) and TF the following formula was applied: $(a_{te}/b)/(c_{te}/d)$, where a_{te} is the total number of bases of TF peaks for a given TE; b is the total number of bases of peaks for a TF; c_{te} is the total number of bases of the te; d is the total number of bases of all TE clusters.

To account for gene co-expression, we first annotated TF peaks over TAIR10 genome using ChIPseeker[112] package in R v1.30 and extracted the ID of downstream genes. We extracted coexpressed gene tables from ATTEDII[93] (version Ath-r.c3-1) and kept the top 10% of both positive and negatively co-expressed genes. For enrichment of TE clusters in chromatin landscapes, chromatin landscapes from Jamge et al.[51] were used. We calculated Fisher's Exact Test comparing enrichment of gene IDs in annotated TEs against chromatin landscapes, both for the alternative hypothesis of being greater or less than expected were calculated. The $-\log_{10}$(p-value) was assigned if the alternative hypothesis of being greater has the lowest p-value, or the $\log_{10}$(p-value) if not.

Acidic extraction of histone proteins in C. merolae

To extract proteins from cells, we collected 1 ml *C. merolae* 10D culture of O.D750 = 1. Cells were centrifuged for 10 minutes at 3000 rcf and washed with 3x with 1 ml ice-cold Nuclei Isolation Buffer (NIB; see Ekwall et al.[129] for component information) without any protease inhibitor, Zn, or Trichostatin A (TSA). We then resuspended the cells in 500 μl of NIB buffer supplemented with 10 ng/μl TSA, 2 mM $ZnSO_4$, and a protease inhibitor cocktail. Next, we transferred the entire mixture into 2 ml tubes and added 0.5 mm Disruptor Beads (cat. SI-BG05, Scientific Industries, USA) until beads reached 1 mm below liquid surface. Next, cells were lysed in a homogenizer (Precellys Evolution, Bertin Technologies, USA) using the following program: 1 minute 8000 rpm, 30 seconds pause on ice, repeated a total of 3 times. We collected the lysate into a fresh tube and centrifuged the sample at 4 °C at maximum speed for 10 minutes and resuspended the pellet in 500 μl of 0.44 N H_2SO_4 before incubating it for 1 hour on ice. Next, we centrifuged the sample at 9,000 rcf for 5 minutes at 4 °C and collected the supernatant to a fresh tube. After repeating the H_2SO_4 extraction on the pellet, we combined the supernatants to a total of 1.0 ml and added 250 μl of 6.1 N TCA to precipitate the proteins for 30 minutes to 1 hour on ice. The final pellet was resuspended in 200 μl 1x Laemmlie buffer in 0.2x PBS and boiled at 95 °C for 5 minutes. The extracted histone proteins were loaded on 15% SDS-gels and were then stained using silver staining.[130]

Coomassie-stained gel bands were cut to 2-3 mm pieces, transferred to 0.6 ml tubes and incubated with different solutions by shaking for 10 minutes at room temperature followed by removal of the supernatant as follows: Gel pieces were washed with 200 μl 100 mM Hepes pH 8.5, destained by 2 repeated rounds of shrinking in 200 μl 50% ACN in 50 mM Hepes and reswelling in 200 μl 100 mM Hepes. Gel pieces were shrunk with 100 μl ACN before being reduced with 100 μl of 1 mg/ml Dithiothreitol in 100 mM Hepes by incubation at 57 °C for 30 minutes and alkylated with 100 μl of 5 mg/ml Iodoacetamide in 100 mM Hepes by incubation at RT for 30 minutes in the dark. Wash steps were repeated as described for destaining with a final shrinking step in ACN. Proteins were either propionylated in the gel before the digest or directly in-gel digested with Trypsin or Lys-C using conditions described below for Trypsin. For propionylation gel pieces were incubated in 30 μl 250 mM N-Succinimidyl propionate (dissolved in ACN) plus 30 μl 100 mM Hepes for 2 hours at RT. The propionylation step was repeated, before the gel pieces were again shrunk with ACN. Gel pieces were incubated in 60 μl of 1M Tris pH 8 for 30 minutes at RT. Gel pieces were washed with 2 rounds of swelling and shrinking as described above using 100 mM Tris instead of Hepes. After the final shrinking, gel pieces were incubated with Trypsin (Promega) at 12.5ng/μl overnight at 37°C. The supernatant containing tryptic peptides was transferred to a fresh tube and gel pieces were extracted by addition of 30 μL 5% formic acid and sonication for 10 minutes in a cooled ultrasonic bath. This step was performed twice. All supernatants were unified. A similar aliquot of each digest was analyzed by LC-MS/MS.

HPLC-MS

The system used was an UltiMate 3000 RSLC nano system coupled to an Orbitrap Exploris 480 mass spectrometer, equipped with a Proxeon nanospray source or to an Orbitrap Eclipse Tribrid mass spectrometer equipped with a FAIMS pro interface and a

Nanospray Flex ion source (all parts Thermo Fisher Scientific). Peptides were loaded onto a trap column (Thermo Fisher Scientific, PepMap C18, 5 mm × 300 μm ID, 5 μm particles, 100 Å pore size) at a flow rate of 25 μL/min using 0.1% TFA as mobile phase. After 10 min, the trap column was switched in line with the analytical column (Thermo Fisher Scientific, PepMap C18, 500 mm × 75 μm ID, 2 μm, 100 Å). Peptides were eluted using a flow rate of 230 nl/min, starting with the mobile phases 98% A (0.1% formic acid in water) and 2% B (80% acetonitrile, 0.1% formic acid) and linearly increasing to 35% B over the next 120min, followed by a gradient to 95% B in 5 min, staying there for 5 min and decreasing in 2 min back to the gradient 98% A and 2% B for equilibration at 30°C.

The Orbitrap Exploris 480 mass spectrometer was operated in data-dependent mode, performing a full scan (m/z range 350-1200, resolution 60,000, normalized AGC target 100%) at 3 different compensation voltages (CV -45, -60, -75), followed each by MS/MS scans of the most abundant ions. MS/MS spectra were acquired using HCD collision energy of 30, isolation width of 1.0 m/z, orbitrap resolution of 30,000, normalized AGC target 200%, minimum intensity of 25,000. Precursor ions selected for fragmentation (including charge state 2-6) were excluded for 20 s. The monoisotopic precursor selection (MIPS) filter and exclude isotopes feature were enabled.

Proteomics data analysis

Raw MS data was loaded into Proteome Discoverer (PD, version 2.5.0.400, Thermo Scientific). All MS/MS spectra were searched using MSAmanda v2.0.0.19924.[114] Trypsin was specified as a proteolytic enzyme cleaving after lysine and arginine (K and R) without proline restriction, allowing for up to 5 missed cleavages. Mass tolerances were set to ±10 ppm at the precursor and ±10 ppm at the fragment mass level. Peptide and protein identification was performed in two steps. An initial search was performed against the *C. merolae* 10D proteome from the UniProt database (Proteome ID: UP000007014; 4,995 protein sequences), with common contaminants appended. Here, carbamidomethylation of cysteine was searched as fixed modification, whereas oxidation of methionine, deamidation of asparagine and glutamine and glutamine to pyro-glutamate conversion at peptide N-termini were defined as variable modifications, as well as propionylation on lysines, serines and threonines and N-termini. Results were filtered for a minimum peptide length of 7 amino acids and 1% FDR at the peptide spectrum match (PSM) and the protein level using the Percolator algorithm[115] integrated in Proteome Discoverer. Additionally, an Amanda score of at least 150 was required. Identified proteins were exported and subjected to a second step search considering phosphorylation of serines, threonines and tyrosines as additional variable modifications, as well as acetylation on lysine (Lys), methylation and dimethylation of Lys and Arg, trimethylation of Lys, and propionylation of methylated Lys. The localization of the post-translational modification sites within the peptides was performed with the tool ptmRS, based on the tool phosphoRS.[116] Identifications were filtered using the filtering criteria described above, including an additional minimum PSM-count per protein in at least one sample of 2. The identifications were subjected to label-free quantification using IMP-apQuant.[117] Proteins were quantified by summing unique and razor peptides and applying intensity-based absolute quantification (iBAQ).[118] Following these procedures, only histone H3 peptides were used in this study.

QUANTIFICATION AND STATISTICAL ANALYSIS

Differential expression analyses on transcriptome data shown in Figures 1A, 1B, 2E, 2G, 4B, 4C, S3A, S3B, S4A, and S4B were performed by using DESeq2 v1.22.2 package in R v4.2.0. TEs and PCGs with $|\log_2 \text{fold change}| > 1$ and adjusted P value < 0.05 were defined as differentially expressed TEs and PCGs. Statistical analysis on the co-expression analysis in the Figure 4F was performed by the hypergeometric test. In the Figure 5D, P values of enrichment of TEs in each chromatin landscape were calculated by using Fisher's Exact Test. In Figures 4E and S1A, enrichment of some TE families was tested by using Fisher's Exact Test. Statistical details of Figures S2E, S3F, and S3G are written in figure legends and corresponding sections in method details.